THE INVARIANCE PRINCIPLE FOR DEPENDENT RANDOM VARIABLES

Patrick Paul Billingsley

1-23

INV - INVARIANCE

VAR - VARIABLE

RAN - RANDOM

THE INVARIANCE PRINCIPLE
FOR DEPENDENT RANDOM
VARIABLES

Letter on cover:

THE INVARIANCE PRINCIPLE FOR
DEPENDENT RANDOM VARIABLES

Patrick Paul Billingsley

[illegible]

25 May 1955

From: LT Patrick P. BILLINGSLEY, USN, 512939/1610
To: Superintendent, U. S. Naval Academy Postgraduate School, Santa Monica, California
Via: Commanding Officer, U. S. Naval Security Station, Washington, D. C.

Subj: Doctoral Dissertation; forwarding of

Ref: (a) U. S. Naval Academy Postgraduate School Regulations

Encl: (1) "The Invariance Principle for Dependent Random Variables," Doctoral Dissertation, two copies

1. From August 1950 to June 1952 I was attached to the U. S. Naval Postgraduate School persuing a course of mathematics at Princeton University, Princeton, New Jersey.

2. On 14 May 1955, I completed the requirements for the degree of Ph. D. in mathematics. In accordance with reference (a), I transmit herewith two copies of my doctoral dissertation.

PATRICK P. BILLINGSLEY

- -

71/wsm
P11/00
Ser 2564

FIRST ENDORSEMENT

25 MAY 1955

From: Commanding Officer, U. S. Naval Security Station, Washington, D. C.
To: Superintendent, U. S. Naval Academy Postgraduate School, Santa Monica, California

1. Forwarded.

T. R. KURTZ, JR.

C. H. HERRICK
By direction

A DISSERTATION

Presented to the

FACULTY OF PRINCETON UNIVERSITY

in Candidacy for the Degree

of Doctor of Philosophy

Recommended for Acceptance by the

Department of

Mathematics

May 1955

The author would like to express his sincere gratitude to Professor William Feller for introducing him to probability theory and for his unfailing help and encouragement during the period of preparation of this work.

P. P. B.

The Invariance Principle for

Dependent Random Variables

By

Patrick Paul Billingsley

(Abstract)

In this paper the Erdős-Kac invariance principle, as generalized by Donsker (Mem. Am. Math. Soc., no. 6 (1951),) is extended to the dependent case. Let C be the space of functions continuous on the closed unit interval, with the uniform topology. If $\{X_n\}$ is a sequence of random variables on a probability space $(\Omega, \mathcal{B}, P)$, let p_n be that element of C which is linear on each of the intervals $((j-1)n^{-1}, jn^{-1})$ and satisfies $p_n(0) = 0$ and $p_n(jn^{-1}) = X_1 + \cdots + X_j$ for $j = 1, \ldots, n$. Thus p_n is a (measurable) mapping of Ω into C. Suppose there exists a sequence $\{a_n\}$ of positive constants such that if the measure P_n is defined by $P_n(A) = P\{a_n^{-1} p_n \in A\}$ for measurable subsets A of C, then $\{P_n\}$ converges weakly to Wiener measure. If this is true we say that the invariance principle holds for $\{X_n\}$. Donsker has shown that the invariance principle holds if $\{X_n\}$ is independent and stationary and X_1 has zero mean and unit variance. In the present paper we prove, after some measure-theoretic preliminaries, that the

invariance principle holds under each of the following conditions. (i) $X_n = f(x_n)$, where f is a function on the state space of a discrete Markov process $\{x_n\}$ satisfying Doeblin's condition. (ii) $\{X_n\}$ is m-dependent. (iii) $\{X_n\}$ is a discrete linear process with m-dependent residuals. (iv) X_n is 1 or 0 according as a recurrent event occurs or not at the nth of a sequence of trials. In each of these four cases the additional assumptions under which the invariance principle is proved are essentially those under which the corresponding central limit theorem has been proved.

CONTENTS

§0. Introduction.

In [9]* Erdős and Kac introduced a new method for proving weak limit theorems for functions of the partial sums of an independent sequence of random variables. Their method consisted in showing first that the limiting distribution is independent of the particular sequence and then computing this distribution for some convenient sequence. Let

$$(0.1) \qquad X_1, X_2, \ldots$$

be an independent sequence of identically distributed random variables with zero means and unit variances. Let $S_k = X_1 + \cdots + X_k$. Erdős and Kac showed that the limiting distribution of

$$(0.2) \qquad n^{-1/2} \max_{k \leq n} S_k$$

(along with several other functions of $S_1, \ldots, S_n$) is independent of the distribution function common to the variables in the sequence $\{X_n\}$. Since the limiting distribution of (0.2) in the Bernoulli case was well known, this argument gave the limiting distribution

---

*Numbers in brackets refer to the bibliography at the end of the paper. The expression "Theorem i.j" refers to the jth theorem of §i, while "Theorem A.j" refers to the jth theorem of the appendix.

of (0.2) under quite general conditions. There followed several papers (cf., e.g., Erdős and Kac [10] and Mark [17]) in which this argument, known as the invariance principle, was applied to various other functions of the partial sums S_k. All of these results were subsequently subsumed under a general theorem due to Donsker [6].

Donsker's result runs essentially as follows. Let C be the space of functions $x(t)$ continuous on the closed unit interval, with the uniform topology, and let W be Wiener measure on C. Let $\{X_n\}$ be a sequence of random variables on some probability measure space $(\Omega, \mathcal{B}, P)$. Let p_n be that element of C which is linear on each of the intervals $((j-1)n^{-1}, jn^{-1})$, $j = 1, \ldots, n$, and satisfies $p_n(jn^{-1}) = S_j$ for $j = 1, \ldots, n$, and $p_n(0) = 0$. That is, let p_n be the random function defined by

$$(0.3) \qquad p_n(t) = S_{j-1} + (nt - j + 1)X_j\,, \quad (j-1)n^{-1} \leq t \leq jn^{-1}, \quad j = 1, \ldots, n,$$

where $S_0 = 0$. Thus p_n is a (measurable) mapping of Ω into C. Let f be any function on C which is continuous in the uniform topology at almost all (W-measure) points of C. Donsker showed that if (0.1) is an independent sequence of random variables which are identically distributed with zero mean and unit variance, then

$$\lim_{n \to \infty} P\{f(n^{-1/2} p_n) \leq a\} = W\{x: f(x) \leq a\}$$

at continuity points a of the function $W\{x: f(x) \leq a\}$. If, for example, $f(x) = \max_{0 \leq t \leq 1} x(t)$, then f satisfies the above conditions and $f(n^{-1/2} p_n) = n^{-1/2} \max_{k \leq n} S_k$, so that

$$\lim_{n \to \infty} P\{n^{-1/2} \max_{k \leq n} S_k \leq a\} = W\{x: \max_{0 \leq t \leq 1} x(t)\} = \begin{cases} 0 & \text{if } a \leq 0 \\ \sqrt{\frac{2}{\pi}} \int_0^a e^{-u^2/2}\, du & \text{if } a \geq 0, \end{cases}$$

where the right hand equality can be established by any one of a number of methods. See [6] for other functions f which lead to interesting limit theorems.

It should be pointed out that in place of the "random polygon" p_n defined by (0.3), Donsker actually worked with the "random step-function" with value S_j throughout the interval $((j-1)n^{-1}, jn^{-1}]$. There is of course no essential difference between the two methods.

There is another way of stating Donsker's result. Let $\mathcal{C}$ be the Borel field generated by the open (uniform topology) subsets of C. Suppose there exists a sequence $\{a_n\}$ of positive constants such that if P_n is a measure defined by setting $P_n(A) = P\{a_n^{-1} p_n \in A\}$ for $A \in \mathcal{C}$, then P_n converges weakly to W. When this is true we say that the <u>invariance principle</u> holds for the sequence $\{X_n\}$ with <u>norming factors</u> a_n. Then

(cf. Theorem 1.1 below) Donsker's result says that the invariance principle holds, with norming factors $n^{1/2}$, provided $\{X_n\}$ is an independent stationary sequence with $E\{X_1\} = 0$ and $E\{X_1^2\} = 1$. The assumption that $\{X_n\}$ is stationary is relatively unimportant. It is the purpose of the present paper to replace the assumption of independence by various weaker hypotheses.

The plan of the paper is as follows. The central results are contained in §§ 4 through 7, those of §§ 1 through 3 being preliminary. These first three sections are devoted to an account of the theory of weak convergence of probability measures on $\mathfrak{C}$ (§ 1), an alternative proof of the existence of Wiener measure (§ 2) and a general invariance principle (§ 3).

In § 3 we sort out those steps in the proof of Donsker's theorem (Theorem 1 of [6]) which depend upon the assumption of independence and state them as the hypotheses of Theorem 3.1. This theorem then gives a set of conditions on the sequence $\{X_n\}$ which insures that the invariance principle holds with a suitable sequence of norming factors. While these conditions are not very pleasing, they can be verified for those sequences $\{X_n\}$ of greatest interest.

Theorems 1.1 and 1.3 are preliminary to § 3. Theorem 1.1 gives several sets of conditions equivalent to weak convergence.

These conditions, with the possible exception of (ii), are well known. Theorem 1.3, on which Theorem 3.1 depends, gives a simple criterion for weak convergence of probability measures on $\mathcal{C}$ in terms of the convergence of the measures of sets of the form

$$\{x:\ \alpha_j \leq x(t) \leq \beta_j,\ (j-1)c^{-1} \leq t \leq jc^{-1},\ j = 1, \ldots, c\},$$

where c is a positive integer and α_j, β_j arbitrary real numbers. This theorem is a slight generalization of one due to Donsker. The proof differs from his in that several arguments "of the Riemann approximation type" are eliminated, which elimination is made possible by condition (ii) of Theorem 1.1. Theorem 1.4, which is essential to the considerations of §4, is the anologue for distributions on C of a well-known limit theorem for distributions on the real line. Its proof depends on condition (ii) of Theorem 1.1.

Theorem 1.2 and §2 are a side issue. Theorem 1.2 is a result, announced by Prohorov [19], on the weak compactness of measures on $\mathcal{C}$. Prohorov has used this theorem to give an elegant proof of the invariance principle in the independent case, but his method seems difficult to apply in those cases to which the present paper is devoted. His result is really an existence theorem, and in §3 we use it to prove the existence of Wiener

measure. This method of proving the existence of a stochastic process is of course not very general (cf. [8, Ch. II] for a general approach) but an alternative proof of this important theorem is interesting.

In §§4 through 7, Theorem 3.1 is specialized in various ways. In §4 the invariance principle is proved for sequences $\{f(x_n)\}$, where f is a function defined on the state space of a discrete Markov process $\{x_n\}$ satisfying Doeblin's hypothesis. The conditions under which this result is obtained are identical with those under which the central limit theorem for such processes is proved in [8].

In §5 we prove the invariance principle for m-dependent sequences of random variables. The best central limit theorems for such sequences are due to Marsaglia [18], and the conditions under which the theorems of §5 are proved are essentially those of his central limit theorems. Donsker's original theorem follows from the results of §5.

§6 treats of discrete linear processes with m-dependent residuals, processes which arise in the analysis of time series. Here we prove the invariance principle under conditions only slightly stronger than those assumed by Diananda [5] in his proof of the central limit theorem for such processes.

Finally, in §7 we prove the invariance principle for the number of occurences of a recurrent event. Here we assume that the recurrence time has a finite second moment.

In the appendix we prove several limit theorems for c-dimensional distribution functions. These theorems are all routine extensions of results well known for the case $c = 1$.

It is doubtful that the results of §§4 through 7 can be substantially improved using present methods, since in each case the invariance principle is proved under conditions virtually the same as those under which the central limit theorem has been proved. It is possible to prove the invariance principle in cases other than those considered here. One can, for example, prove it for martingales, as Lévy [16] has the central limit theorem, or under the assumptions of Bernstein's "lemme fondamental" [2]. Although no applications have been essayed, the cases treated in §§4 through 7 are those of greatest interest for the applications.

§1. Convergence of measures on the space of continuous functions.

In this section we prove two useful theorems on the convergence of probability measures on the space of continuous functions.

Consider first an arbitrary metric space $\mathfrak{X}$ with metric ρ. In what follows we will be interested in the cases in which $\mathfrak{X}$ is either the space of continuous functions or a Euclidean space. Let $\mathfrak{B}$ be the collection of Borel sets, that is, the Borel field generated by the open sets. If P_n, P are probability measures on $\mathfrak{B}$, we say that P_n converges weakly to P (in symbols, $P_n \Rightarrow P$) if

$$\int_{\mathfrak{X}} f \, dP_n \longrightarrow \int_{\mathfrak{X}} f \, dP$$

for all bounded continuous functions f.

Theorem 1.1 gives several convenient sets of conditions equivalent to weak convergence. For its proof we require the following variation on Urysohn's lemma.

Lemma 1.1. If A and B are closed sets with $\rho(A, B) > 0$, then there exists a function $f(x)$ which is 1 on A, 0 on B, everywhere between 0 and 1, and uniformly continuous on $\mathfrak{X}$.

Proof: We may of course assume that A and B are non-empty. With the exception of uniform continuity, it is clear that the function

$$f(x) = \frac{\rho(x, B)}{\rho(x, B) + \rho(x, A)}$$

has the required properties. To prove that f is uniformly continuous observe first that (cf. [1, p.57])

$$|\rho(x, A) - \rho(y, A)| \leq \rho(x, y)$$

and

$$\rho(x, A) + \rho(x, B) \geq \rho(A, B) .$$

From these inequalities it follows that

$$|f(x) - f(y)| \leq \frac{|\rho(x, B) - \rho(y, B)|}{\rho(x, B) + \rho(x, A)}$$

$$+ \rho(y, B) \left| \frac{1}{\rho(x, B) + \rho(x, A)} - \frac{1}{\rho(y, B) + \rho(y, A)} \right|$$

$$\leq \frac{\rho(x, y)}{\rho(A, B)} + \frac{\rho(y, A)}{\rho(y, B) + \rho(y, A)} \; \frac{|\rho(y, B) - \rho(x, B)| + |\rho(y, A) - \rho(x, A)|}{\rho(x, B) + \rho(x, A)}$$

$$\leq \frac{3}{\rho(A, B)} \rho(x, y) .$$

Hence f is uniformly continuous on $\mathfrak{X}$.

In what follows we denote the closure, interior and boundary of a set A by $\bar{A}$, A° and $\tilde{A}$, respectively. If P is a probability

measure on $\mathfrak{X}$ and f is a measurable function then $P\{x: f(x) \leq a\}$ is a function of a which we call the P-distribution function of f .

Theorem 1.1. The following statements are equivalent.

(i) $P_n \Longrightarrow P$.

(ii) $\int_{\mathfrak{X}} f \, dP_n \to \int_{\mathfrak{X}} f \, dP$ for bounded uniformly continuous functions f .

(iii) $P(A) \geq \limsup_{n \to \infty} P_n(A)$ for closed sets A .

(iv) $P(A) = \lim_{n \to \infty} P_n(A)$ for sets $A \in \mathcal{B}$ such that $P(\tilde{A}) = 0$.

(v) For any function f which is continuous except on a set of P-measure zero, the P_n-distribution function of f converges to the P-distribution function of f at each continuity point of the latter.

Proof. We will prove in turn the implications (i) $\to$ (ii) $\to$ (iii) $\to$ (iv) $\to$ (v) $\to$ (i) . The implication (i) $\to$ (ii) is trivial.

(ii) $\to$ (iii) . Suppose that A is closed and $\varepsilon > 0$ given. We may assume that A is neither the empty set nor the whole space. For $\delta > 0$ let $U_\delta = \{x: \rho(x, A) < \delta\}$. Then U_δ is open and $U_\delta \downarrow A$ as $\delta \downarrow 0$, since A is closed. Hence there exists a δ such that $P(U_\delta - A) < \varepsilon$. Clearly $\rho(A, \mathfrak{X} - U_\delta) \geq \delta > 0$.

Therefore, by Lemma 1.1, there exists a uniformly continuous function f which is 1 on A, 0 on $\mathfrak{X} - U_\delta$ and everywhere between 0 and 1. Now

$$\int_{\mathfrak{X}} f \, dP_n \longrightarrow \int_{\mathfrak{X}} f \, dP$$

by (ii), and

$$\int_{\mathfrak{X}} f \, dP_n \geq P_n(A),$$

while

$$\int_{\mathfrak{X}} f \ dP \leq P(A) + P(U_\delta - A) \leq P(A) + \varepsilon \ .$$

From these three relations it follows that

$$\limsup_{n \to \infty} P_n(A) \leq P(A) + \varepsilon \ .$$

Since ε is arbitrary, (iii) follows.

<u>(iii) → (iv)</u>. Suppose that $P(\tilde{A}) = 0$. Then

$$P(A) = P(\bar{A}) \geq \limsup_{n \to \infty} P_n(\bar{A}) \geq \limsup_{n \to \infty} P_n(A) \ . \tag{1.1}$$

Since the boundary of $\mathfrak{X} - A$ also has P-measure zero we have in the same way,

$$P(\mathfrak{X} - A) \geq \limsup_{n \to \infty} P_n(\mathfrak{X} - A) \ . \tag{1.2}$$

But (1.1) and (1.2) imply

$$P(A) = \lim_{n \to \infty} P_n(A) \ .$$

<u>(iv) → (v)</u>. Let F_n and F be respectively the P_n-distribution functions and the P-distribution function of f and let A be the set of points at which f is discontinuous. Then

$$\overline{\{x: f(x) \leq a\}} \subset \{x: f(x) \leq a\} \cup A$$

and

$$\{x: f(x) < a\} - A \subset \{x: f(x) \leq a\}^{\circ} \ ,$$

so that the boundary of $\{x: f(x) \leq a\}$ is contained in $\{x: f(x) = a\} \cup A$. Since $P(A) = 0$, $F(a) = F(a-0)$ implies that the boundary of $\{x: f(x) \leq a\}$ has P-measure zero, and hence that $F_n(a) \to F(a)$.

(v) → (i). Let F_n and F be the P_n-distribution functions and the P-distribution function of the bounded continuous function f. We assume that $F_n(a) \to F(a)$ if F is continuous at a and must show that

$$\int_{\mathfrak{X}} f \, dP_n \longrightarrow \int_{\mathfrak{X}} f \, dP .$$

But this last statement is equivalent to

$$\int_{-M}^{M} a \, dF_n(a) \longrightarrow \int_{-M}^{M} a \, dF(a),$$

where M is the bound of f. But (1.3) is easy to establish (cf. [3, p. 74]). This completes the proof of Theorem 1.1.

We note at this point the well-known fact that if $\mathfrak{X}$ is c-dimensional Euclidean space then equivalent to each of the conditions of Theorem 1.1 is the condition that if F_n and F are the distribution functions corresponding to P_n and P respectively, then

$$\lim_{n \to \infty} F_n(a_1, \ldots, a_c) = F(a_1, \ldots, a_c)$$

at each point $(a_1, \ldots, a_c)$ such that

[illegible]

[illegible]

$$[illegible]$$

[illegible]

$$[illegible]$$

[illegible]

[illegible]

$$[illegible]$$

[illegible]

$$F(\alpha_1, \ldots, \alpha_c) = \sup_{\alpha_i > \beta_i} F(\beta_1, \ldots, \beta_c),$$

i.e., at continuity points of F.

Let C be the space of functions $x(t)$ continuous on the closed unit interval, with the metric

$$\rho(x, y) = \sup_{0 \le t \le 1} |x(t) - y(t)| .$$

Then C is a complete, separable metric space. Let $\mathcal{C}$ be the collection of Borel sets. The Borel field $\mathcal{C}$ is generated by the sets of the form $\{x: x(t) \le \alpha\}$. That such sets belong to $\mathcal{C}$ is obvious, and to see that they generate $\mathcal{C}$ it is enough to observe that

$$\{x: \rho(x, x_0) \le \delta\} = \bigcap \{x: |x(r) - x_0(r)| \le \delta\},$$

where the intersection extends over all rationals r in the unit interval.

If $t_1, \ldots, t_k$ are fixed points in the closed unit interval, $[x(t_1), \ldots, x(t_k)]$ defines, as x varies over C, a k-dimensional random vector on C which we denote $[x_{t_1}, \ldots, x_{t_k}]$.

The first of the two theorems concerning the convergence of probability measures on C which we will need is due to Prohorov [19].

Theorem 1.2. Suppose that $\{P_n\}$ is a sequence of probability measures on $\mathcal{C}$ with the property that for each $\varepsilon > 0$ there exists a compact set K_ε such that $P_n(K_\varepsilon) > 1 - \varepsilon$ for all n. Then there exists a sequence $\{n_\nu\}$ and a probability measure P such that $P_{n_\nu} \Rightarrow P$ as $\nu \to \infty$.

Proof. Let $\{r_k, k \geq 1\}$ be an ordering of the rationals of the closed unit interval. For each n and k

$$(1.4) \qquad \mu_{k,n}(S) = P_n\{[x_{r_1}, \ldots, x_{r_k}] \in S\}$$

is a probability measure of k-dimensional Borel sets S. Let $F_{k,n}$ be the distribution function corresponding to $\mu_{k,n}$. For each k it is possible by Helly's theorem to find an increasing sequence $\{n_\nu\}$ of integers and a function $F_k(\alpha_1, \ldots, \alpha_c)$ such that F_k is everywhere between 0 and 1, is non-decreasing in each variable, is continuous from above and

$$(1.5) \qquad \lim_{\nu \to \infty} F_{k,n_\nu}(\alpha_1, \ldots, \alpha_k) = F_k(\alpha_1, \ldots, \alpha_k)$$

at continuity points of F_k. By the diagonal method it is possible to choose a single sequence $\{n_\nu\}$ so that (1.5) holds for all k simultaneously.

For a fixed k let

$$S_\varepsilon = \{[x(r_1), \ldots, x(r_k)]: \ x \in K_\varepsilon\}.$$

Then S_ε is a k-dimensional Borel set with

$$\mu_{k,n}(S_\varepsilon) > 1 - \varepsilon$$

for all n. Since K_ε is compact, S_ε is bounded. From these facts it follows that F_k must have total variation 1, i.e., that it is the distribution function corresponding to some probability measure μ_k. And from the remark following the proof of Theorem 1.1 we conclude that

$$(1.6) \qquad \mu_{k,n_\nu} \Rightarrow \mu_k \qquad (\nu \to \infty)$$

for all k.

We now use the measures μ_k to set up a measure on C in a way similar to that used by Kolmogorov [15] in his fundamental existence theorem. His theorem as such is not applicable here since we are working in the space of continuous functions rather than the space of all functions. Let $\mathcal{F}$ be the (finitely additive) field of sets of the form

$$(1.7) \qquad A = \{x: [x(r_1), \ldots, x(r_k)] \in S\},$$

where k is any integer and S is a k-dimensional Borel set. For such a set A put $P(A) = \mu_k(S)$. Since there are other representations of A, we must show that this definition is consistent. Suppose then that in addition to (1.7) we have

$$(1.8) \qquad A = \{x: [x(r_1), \ldots, x(r_j)] \in S'\},$$

where $j > k$ and S' is a j-dimensional Borel set. From the fact that for any point $(\zeta_1, \ldots, \zeta_k)$ of k-space there exists an $x \in C$ with $[x(r_1), \ldots, x(r_k)] = (\zeta_1, \ldots, \zeta_k)$ it follows that $S' = \{(\zeta_1, \ldots, \zeta_j): (\zeta_1, \ldots, \zeta_k) \in S\}$. From this and the definition (1.4) we have

$$\mu_{k,n}(S) = \mu_{j,n}\{(\zeta_1, \ldots, \zeta_j): (\zeta_1, \ldots, \zeta_k) \in S\}$$

for all n. Hence by (1.6)

$$(1.9) \qquad \mu_k(S) = \mu_j\{(\zeta_1, \ldots, \zeta_j): (\zeta_1, \ldots, \zeta_k) \in S\},$$

provided the boundary of S has μ_k-measure zero. But this clearly implies (1.9) for all S, which establishes the consistency of the definition of $P(A)$.

It is easy to show that P is a finitely additive measure on $\mathcal{F}$ and that $P(C) = 1$. We now prove that P is completely additive on $\mathcal{F}$. Suppose then that $\{A_k\}$ is a non-increasing sequence of $\mathcal{F}$ sets with $P(A_k) \geq L > 0$ for all k. We will show that the A_k have a non-empty intersection. For notational convenience we assume that A_k is defined in terms of the first k of the $\{r_i\}$:

$$A_k = \{x: [x(r_1), \ldots, x(r_k)] \in S_k\}.$$

Choose ε so that $0 < 2\varepsilon < L$ and let K_ε be as in the hypothesis of the theorem. Let $U_k \subset S_k$ be a compact k-dimensional Borel set such that $\mu_k(S_k - U_k) < \varepsilon 2^{-k}$. Now let $V_k = \{x: [x(r_1), \ldots, x(r_k)] \in U_k\}$, let $W_k = V_1 \cdots V_k$ and let $Z_k = \{[x(r_1), \ldots, x(r_k)]: x \in W_k\}$. Then

(1.10) $$\mu_k(S_k - Z_k) = P(A_k - W_k) \leq \varepsilon \quad .$$

Finally, let $J_k = \{[x(r_1), \ldots, x(r_k)] : x \in K_\varepsilon\}$. Now J_k is compact and hence closed. In fact, if $\zeta_m \in J_k$ for $m \geq 1$, one can select $x_m \in K_\varepsilon$ such that $[x_m(r_1), \ldots, x_m(r_k)] = \zeta_m$. Since K_ε is compact there exists a sequence $\{m_i\}$ of integers and an $x \in K_\varepsilon$ such that $\lim_{i\to\infty} x_{m_i} = x$. But then $\lim_{i\to\infty} \zeta_{m_i} = [x(r_1), \ldots, x(r_k)] \in J_k$, so that J_k contains a limit point of the sequence $\{\zeta_m\}$. Since J_k is closed we have by (1.6),

(1.11) $$\mu_k(J_k) \geq \limsup_{\nu\to\infty} \mu_{k,n_\nu}(J_k) \geq \limsup_{\nu\to\infty} P_{n_\nu}(K_\varepsilon) \geq 1 - \varepsilon \quad .$$

From (1.10) and (1.11) it follows that

$$\mu_k(Z_k J_k) \geq \mu_k(S_k) - 2\varepsilon \geq L - 2\varepsilon > 0 \; .$$

Hence $Z_k J_k$ is non-empty, which implies that $W_k K_\varepsilon$ is non-empty. Since the W_k form a non-increasing sequence of sets, and each is obviously closed, the $W_k K_\varepsilon$ have a non-empty intersection by the compactness of K_ε. Since $W_k \subset A_k$, the A_k have a non-empty

intersection. Hence P is completely additive on $\mathcal{F}$.

Since $\mathcal{F}$ generates $\mathcal{C}$, P can be extended to a probability measure on $\mathcal{C}$. We now complete the proof of the theorem by showing that $P_{n_\nu} \Rightarrow P$. By Theorem 1.1 it suffices to prove that

$$(1.12) \qquad P(A) \geq \limsup_{\nu \to \infty} P_{n_\nu}(A)$$

for closed sets A. For a given $\varepsilon > 0$ let

$$S_k = \{[x(r_1), \ldots, x(r_k)] : x \in AK_\varepsilon\}.$$

Now S_k is closed for the same reason the set J_k above was. If $B_k = \{x: [x(r_1), \ldots, x(r_k)] \in S_k\}$, it is clear that $\{B_k\}$ is a non-increasing sequence of sets of $\mathcal{C}$ with $AK_\varepsilon \subset \bigcap_k B_k$. Suppose on the other hand that $x \in \bigcap_k B_k$. From the definition of S_k and B_k it is possible to find a sequence $\{y_m\}$ of points of AK_ε such that $y_m(r_j) = x(r_j)$ for $j = 1, \ldots, k$. Since AK_ε is compact there exists a sequence $\{m_i\}$ of integers and a point $y \in AK_\varepsilon$ with $\lim_{i \to \infty} y_{m_i} = y$. But then $x = y$ by continuity, and $x \in AK_\varepsilon$. Hence $AK_\varepsilon = \bigcap_k B_k$. It is therefore possible to choose k so that

$$(1.13) \qquad \varepsilon + P(AK_\varepsilon) \geq P(B_k) = \mu_k(S_k).$$

Since S_k is closed we have by (1.6) and Theorem 1.1,

$$(1.14) \qquad \mu_k(S_k) \geq \limsup_{\nu \to \infty} \mu_{k,n_\nu}(S_k).$$

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

Now by the hypothesis of the theorem,

$$\mu_{k,n_\nu}(S_k) \geq P_{n_\nu}(AK_\varepsilon) \geq P_{n_\nu}(A) - \varepsilon$$

for all ν . Hence

$$\limsup_{\nu \to \infty} \mu_{k,n_\nu}(S_k) \geq \limsup_{\nu \to \infty} P_{n_\nu}(A) - \varepsilon \quad . \tag{1.15}$$

But now (1.12) follows from (1.13), (1.14) and (1.15), since ε is arbitrary. This completes the proof of Theorem 1.2 .

In order to apply this theorem we need a criterion for compactness of sets in C . The following standard lemma gives a convenient criterion in terms of the modulus of continuity, which for our purposes is best defined by

$$M(x, \delta) = \sup \{ |x(s) - x(t)| \; : \; s, t \in [0, 1], \; |s - t| \leq \delta \}$$

for $x \in C$ and $\delta > 0$.

Lemma 1.2 . If A is a closed set such that

$$\sup_{x \in A} |x(0)| < \infty \tag{1.16}$$

and

$$\lim_{\delta \to 0} \sup_{x \in A} M(x, \delta) = 0 , \tag{1.17}$$

then A is compact.

<u>Proof</u>. It is in the first place easy to show that

$$(1.18) \qquad \sup_{x \in A} |x(t)| < \infty$$

for each t in the unit interval. Suppose $\{x_n\} \subset A$. By (1.18) and the diagonal procedure it is possible to find a sequence $\{n_\nu\}$ of integers such that

$$\lim_{\nu \to \infty} x_{n_\nu}(r) = L(r)$$

exists for each rational r in the unit interval. We prove that $x_{n_\nu}(t)$ is uniformly fundamental. Given $\varepsilon > 0$, choose δ_0 so that $M(x, \delta) < \varepsilon$ if $\delta \leq \delta_0$ and $x \in A$. Then choose rational $r_0, \ldots, r_k$ so that

$$0 = r_0 < r_1 < \ldots < r_k = 1$$

and $r_i - r_{i-1} < \delta_0$ for $i = 1, \ldots, k$. Finally, choose N so that

$$|x_{n_\lambda}(r_i) - x_{n_\nu}(r_i)| < \varepsilon$$

for $i = 0, \ldots, k$, provided $\lambda, \nu > N$. It follows immediately that

$$|x_{n_\lambda}(t) - x_{n_\nu}(t)| < 3\varepsilon$$

for all $t \in [0, 1]$, provided $\lambda, \nu > N$. Thus $x_{n_\nu}(t)$ is uniformly fundamental and hence converges uniformly to an element of C, which must lie in A. Therefore A is compact.

[illegible]

[illegible]

$$[illegible]$$

[illegible]

$$[illegible]$$

[illegible]

$$[illegible]$$

[illegible]

$$[illegible]$$

[illegible]

$$[illegible]$$

[illegible]

In order to prove the second of the convergence theorems of this section we need a lemma which is a slight variation on a result due to Donsker [6]. If $x \in C$ and c is a positive integer, define, for $j = 1, \ldots, c$,

$$a_j(x) = \inf \{x(t) : (j-1)c^{-1} \leq t \leq jc^{-1}\}$$

$$b_j(x) = \sup \{x(t) : (j-1)c^{-1} \leq t \leq jc^{-1}\} \quad .$$

Now let

$$\pi_c(x) = (a_1(x), \ldots, a_c(x), b_1(x), \ldots, b_c(x)) .$$

Thus π_c maps C into $2c$-space. It is easy to show that π_c is continuous. If φ is bounded and continuous on $2c$-space, then $\varphi(\pi_c(x))$ is a bounded continuous function on C. Let $\mathfrak{A}$ be the set of functions arising in this way (as c and φ vary).

<u>Lemma 1.3</u>. Let f be a bounded uniformly continuous function on C. Then there exists a pair of sequences of functions $\{f'_c\}$ and $\{f''_c\}$, all belonging to $\mathfrak{A}$, such that for all c and x,

(1.19) $$f'_c(x) \leq f(x) \leq f''_c(x) ,$$

such that f'_c and f''_c are uniformly bounded, such that

(1.20) $$\lim_{c \to \infty} (f''_c(x) - f'_c(x)) = 0$$

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

for all $x \in C$.

Proof. For each c let M_x^c be the (non-empty) set of $y \in C$ such that $a_j(x) \leq y(t) \leq b_j(x)$ for $(j-1)c^{-1} \leq t \leq jc^{-1}$, $j = 1, \ldots, c$. And define

$$f'_c(x) = \inf \{ f(y) : y \in M_x^c \}$$

$$f''_c(x) = \sup \{ f(y) : y \in M_x^c \} \quad .$$

It is clear that f'_c and f''_c satisfy (1.19) and that they are bounded by the bound of f . And (1.20) follows from the uniform continuity of f .

There remains only the proof that f'_c and f''_c belong to $\mathfrak{A}$. We consider only the case of f''_c . Let S be the set of points $(\zeta_1, \ldots, \zeta_{2c})$ of $2c$-space having the property that

$$(1.21) \qquad \zeta_j \leq \zeta_{c+j} \quad , \qquad\qquad j = 1, \ldots, c ,$$

and

$$(1.22) \qquad [\zeta_j, \zeta_{c+j}] \cap [\zeta_{j+1}, \zeta_{c+j+1}] \neq 0 , \quad j = 1, \ldots, c-1 .$$

Then S is obviously closed. If $(\zeta_1, \ldots, \zeta_{2c}) \in S$ define

$$\varphi(\zeta_1, \ldots, \zeta_{2c}) = \sup \{ f(y) : \zeta_j \leq y(t) \leq \zeta_{c+j}, (j-1)c^{-1} \leq t \leq jc^{-1} , \ j = 1, \ldots, c \} ,$$

[illegible]

where the set over which the supremum is extended is non-empty by (1.21) and (1.22). Obviously $\pi_c(x) \in S$ for all $x \in C$, $\varphi(\pi_c(x)) = f''_c(x)$, and φ is bounded. Suppose we prove that φ is continuous on S. Then it is possible by Urysohn's extension theorem [1, p. 73] to extend φ to all of 2c-space in such a way that it remains bounded and continuous. Hence the proof will be completed if we show φ is continuous on S.

Suppose $(\zeta_1, \ldots, \zeta_{2c}) \in S$ and that $\varepsilon > 0$ is given. We will find a $\delta > 0$ such that

$$(1.23) \qquad |\varphi(\zeta_1, \ldots, \zeta_{2c}) - \varphi(\xi_1, \ldots, \xi_{2c})| < \varepsilon$$

provided

$$(1.24) \qquad |\zeta_j - \xi_j| < \delta \;, \qquad\qquad j = 1, \ldots, 2c \,,$$

and $(\xi_1, \ldots, \xi_{2c}) \in S$. By the uniform continuity of f there exists a δ such that $|f(x) - f(y)| < \varepsilon/2$ if $\rho(x, y) < \delta$. Suppose now that $(\xi_1, \ldots, \xi_{2c})$ satisfies the above conditions. It is clearly possible to find x and y in C such that for $j = 1, \ldots, c$,

$$\zeta_j = \inf\{x(t) : (j-1)c^{-1} \le t \le jc^{-1}\} \,,$$

$$\zeta_{c+j} = \sup\{x(t) : (j-1)c^{-1} \le t \le jc^{-1}\} \,,$$

$$\xi_j = \inf\{y(t) : (j-1)c^{-1} \le t \le jc^{-1}\} \,,$$

$$\xi_{c+j} = \sup\{y(t) : (j-1)c^{-1} \le t \le jc^{-1}\} \,.$$

Then

$$\varphi(\zeta_1, \ldots, \zeta_{2c}) = \varphi(\pi_c(x)) = f''_c(x),$$

$$\varphi(\xi_1, \ldots, \xi_{2c}) = \varphi(\pi_c(y)) = f''_c(y),$$

and it suffices to show that

$$(1.25) \qquad |f''_c(x) - f''_c(y)| < \varepsilon \qquad .$$

By the definition of f''_c there exists a $z \in M^c_x$ such that

$$(1.26) \qquad f''_c(x) \leq f(z) + \varepsilon/2 \ .$$

Now let $z'(t) = z(t)$ at points t where

$$(j-1)c^{-1} \leq t \leq jc^{-1} \quad , \qquad \xi_j \leq z(t) \leq \xi_{c+j} \ ,$$

let $z'(t) = \xi_j$ where

$$(j-1)c^{-1} \leq t \leq jc^{-1} \quad , \qquad z(t) < \xi_j \quad ,$$

and let $z'(t) = \xi_{c+j}$ where

$$(j-1)c^{-1} \leq t \leq jc^{-1} \quad , \qquad \xi_{c+j} < z(t) \ .$$

It follows that $z' \in M^c_y$, so that

$$(1.27) \qquad f(z') \leq f''_c(y) \ ,$$

and from (1.24) it follows that $\rho(z, z') < \delta$, and hence by the

choice of δ ,

(1.28) $\quad f(z) \leq f(z') + \varepsilon/2$.

From (1.26), (1.27) and (1.28) we have

$$f_c''(x) - f_c''(y) < \varepsilon \ .$$

The symmetric inequality follows in the same way, and (1.25) results. This completes the proof of the lemma.

We come now to the second of the two convergence theorems of this section. For any integer c and real numbers $\alpha_1, \ldots, \alpha_c$, $\beta_1, \ldots, \beta_c$, consider the set

(1.29) $\quad E = \{x : \alpha_j \leq x(t) \leq \beta_j, \ (j-1)c^{-1} \leq t \leq jc^{-1}, \ j = 1, \ldots, c\}$.

Theorem 1.3. Suppose that for probability measures P_n and P on $\mathcal{C}$ we have $P_n(E) \to P(E)$ for all sets E of the form (1.29) for which $P(\tilde{E}) = 0$. Then $P_n \Rightarrow P$.

Proof. We show first that

(1.30) $\quad \int_C f \, dP_n \longrightarrow \int_C f \, dP$

for any function f in $\mathfrak{A}$. For a fixed integer c define

$$\mu_n(S) = P_n(\pi_c^{-1}(S)) \ ,$$

$$\mu(S) = P(\pi_c^{-1}(S)) \ ,$$

for 2c-dimensional Borel sets S . Now if

$$S = \{ (\zeta_1, \ldots, \zeta_{2c}) : \zeta_j \geq a_j, \ \zeta_{c+j} \leq \beta_j, \ j = 1, \ldots, c \} ,$$

then $\pi_c^{-1}(S) = E$, so that $\mu_n(S) \to \mu(S)$, provided $\mu(\tilde{S}) = 0$. But this obviously implies $\mu_n \Rightarrow \mu$. Hence

$$(1.31) \qquad \int_{R^{2c}} \varphi \, d\mu_n \longrightarrow \int_{R^{2c}} \varphi \, d\mu$$

for any bounded continuous function φ . But if $f(x) = \varphi(\pi_c(x))$, (1.30) follows from (1.31) by a transformation of the integrals involved. Hence (1.30) holds for any function f of $\mathfrak{A}$.

Now suppose f is a bounded uniformly continuous function on C . By Lemma 1.3 there exists a pair of sequences $\{f'_c\}$ and $\{f''_c\}$ of functions of $\mathfrak{A}$ which are uniformly bounded and satisfy (1.19) and (1.20). By (1.20) and uniform boundedness we have

$$(1.32) \qquad \lim_{c \to \infty} \int_C (f''_c - f'_c) \, dP = 0 .$$

But from (1.32) and (1.19) it easily follows that (1.30) holds. And since (1.30) holds for any bounded uniformly continuous function f , we have $P_n \Rightarrow P$.

In the following sections we will be concerned with "random polygons". For the purposes of this discussion we define a polygon to be an element p of C with the property that the unit

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

$$[illegible]$$

[illegible]

[illegible]

$$[illegible]$$

[illegible]

[illegible]

interval can be decomposed into a finite number of sub-intervals over each of which p is linear. Suppose $(\Omega, \mathcal{B}, P)$ is a probability measure space on which random variables $X_0, X_1, \ldots, X_n$ are defined. Associated with each point ω of Ω we define a polygon $p = p_\omega$ by the equations

$$p_\omega(t) = (j - nt)X_{j-1}(\omega) + (nt - j + 1)X_j(\omega), \quad (j-1)n^{-1} \leq t \leq jn^{-1}, \quad j = 1, \ldots, n, \tag{1.33}$$

that is, p is the polygon with vertices at the points (jn^{-1}, X_j). We will be interested in the distribution properties of p, that is, we will be interested in probabilities $P\{p \in A\}$, where $A \in \mathcal{C}$. In order to show that $P\{p \in A\}$ has meaning we must show that $\{\omega : p_\omega \in A\} \in \mathcal{B}$, i.e., that the mapping $\omega \longrightarrow p_\omega$ is measurable. Since the collection of A's for which this holds forms a Borel field, it suffices to prove it for A's of the form $A = \{x : x(t) \leq \alpha\}$. But if A has this form, and $(j-1)n^{-1} \leq t \leq jn^{-1}$, then by (1.33)

$$\{\omega : p_\omega \in A\} = \{\omega : (j - nt)X_{j-1}(\omega) + (nt - j + 1)X_j(\omega) \leq \alpha\},$$

which lies in $\mathcal{B}$ since the X_j are measurable.

We can use this result to set up on $\mathcal{C}$ a probability measure P' which gives unit mass to the set of polygons which are linear on each of the intervals $((j-1)n^{-1}, jn^{-1})$, $j = 1, \ldots, n$, and under

which the random vector $(x_0, x_{n-1}, x_{2n-1}, \ldots, x_1)$ has a prescribed distribution function. In fact if $X_0, X_1, \ldots, X_n$ are random vectors (on some probability space $(\Omega, \mathcal{B}, P)$) having the prescribed distribution function, and if P' is defined by $P'(A) = P\{p \in A\}$ for $A \in \mathcal{C}$, then P' obviously has the desired properties.

A useful fact in the theory of distributions on the real line is that if the distributions of a sequence $\{X_n\}$ of random variables converge weakly to F, then so do the distributions of $\{X_n + Y_n\}$, provided $\operatorname{p\,lim}_{n \to \infty} Y_n = 0$. We conclude this section with a theorem which plays an analogous role in the theory of distributions on C. The theorem and its proof obviously remain unchanged if C is replaced by any Banach space.

Let $\{X_n\}$ and $\{Y_n\}$ be two sequences of measurable functions on some probability measure space $(\Omega, \mathcal{B}, P)$, with values in C. That is, we assume that $X_n(\omega) \in C$ for $\omega \in \Omega$ and that $\{\omega : X_n(\omega) \in A\} \in \mathcal{B}$ if $A \in \mathcal{C}$, and similarly for Y_n. Let P_n, P_n' and P_n'' be the distributions on C of X_n, Y_n and $X_n + Y_n$ respectively:

$$P_n(A) = P\{X_n \in A\},$$

$$P_n'(A) = P\{Y_n \in A\},$$

$$P_n''(A) = P\{X_n + Y_n \in A\},$$

for $A \in \mathfrak{C}$. Let U be the measure on $\mathfrak{C}$ which places unit mass at the function which is identically zero.

Theorem 1.4. If $P_n \Rightarrow Q$ and $P'_n \Rightarrow U$ then $P''_n \Rightarrow Q$.

Proof. It is clear that $P'_n \Rightarrow U$ is equivalent to the statement that

(1.34) $$\lim_{n \to \infty} P\{\rho(0, Y_n) \geq \varepsilon\} = 0$$

for all $\varepsilon > 0$. Let f be a bounded uniformly continuous function on C. Given ε choose δ so that $|f(x) - f(y)| < \varepsilon$ if $\rho(x, y) < \delta$. Then

$$P\{|f(X_n) - f(X_n + Y_n)| \geq \varepsilon\} \leq P\{\rho(X_n, X_n + Y_n) \geq \delta\} \to 0$$

by (1.34). Hence

(1.35) $$\underset{n \to \infty}{p \lim} (f(X_n) - f(X_n + Y_n)) = 0 .$$

Now by Theorem 1.1, $P\{f(X_n) \leq a\} \to Q\{x : f(x) \leq a\}$ at continuity points of the later function. Hence by (1.35) and the above-mentioned fact in the theory of distributions on the real line, $P\{f(X_n + Y_n) < a\} \to Q\{x : f(x) \leq a\}$. Since f is bounded this implies

$$\int_C f \, dP''_n \longrightarrow \int_C f \, dQ .$$

Hence by Theorem 1.1 ((ii)→(i)) we have $P''_n \Rightarrow Q$.

§2. The existence of Wiener measure.

A fundamental problem connected with the measure-theoretic aspect of probability theory is that of proving the existence of stochastic processes having specified properties. In this section we give a proof, based on Theorem 1.2, of the existence of the Wiener process.

A probability measure P on $\mathcal{C}$ is called <u>Gaussian</u> if the P-distribution of the random vector $(x_{t_1}, \ldots, x_{t_n})$ is, for any set $(t_1, \ldots, t_n)$ of points of the unit interval, a normal distribution with zero means. For such a measure P the covariance function

$$R(s,t) = \int_C x_s x_t \, dP$$

is defined for all s and t in the unit interval. Moreover R completely determines P. In fact R determines the covariance matrix and hence the distribution of each $(x_{t_1}, \ldots, x_{t_n})$, and these in turn clearly determine P. The question is, given R, does there exist a Gaussian measure P with R as its covariance function? The following theorem gives an affirmative answer in the case $R(s,t) = \min(s,t)$. The Gaussian measure having this covariance function is called <u>Wiener measure</u>, and is denoted here by W. It is a simple matter to show that if

$t_1 < t_2 < \cdots < t_n$, then the random variables

$$x_{t_2} - x_{t_1},\ x_{t_3} - x_{t_2},\ \ldots,\ x_{t_n} - x_{t_{n-1}},$$

are, under W, normally and independently distributed with zero means and variances

$$t_2 - t_1,\ t_3 - t_2,\ \ldots,\ t_n - t_{n-1}.$$

<u>Theorem 2.1.</u> There exists a Gaussian measure W such that

$$\int_C x_s x_t \, dW = \min(s,t) \tag{2.1}$$

for all s and t in the closed unit interval.

<u>Proof.</u> For each positive integer n let $A_n \in \mathfrak{C}$ be the set of polygons which are linear on each of the intervals $((j-1)2^{-n}, j2^{-n})$, $j = 1, \ldots, 2^n$. Let P_n be a probability measure on $\mathfrak{C}$ such that $P_n(A_n) = 1$ and such that the P_n-distribution of

$$(x_0, x_{1\cdot 2^{-n}}, x_{2\cdot 2^{-n}}, \ldots, x_1) \tag{2.2}$$

is normal with zero means and covariance matrix (λ_{ij}), where $\lambda_{ij} = 2^{-n}\min(i,j)$. That such a P_n exists follows from the remarks preceding Theorem 1.4 and the fact that the matrix (λ_{ij}) is positive semi-definite (in fact

$\Sigma_{i,j=0}^{2^n} \lambda_{ij} u_i u_j = 2^{-n} \Sigma_{i=1}^{2^n} (\Sigma_{j=i}^{2^n} u_j)^2 \geq 0$.) Note that under P_n the differences

$$(2.3) \qquad x_{1\cdot 2^{-n}} - x_0 , \; x_{2\cdot 2^{-n}} - x_{1\cdot 2^{-n}} , \; \ldots$$

are independently and normally distributed, each with variance 2^{-n} .

We will first show that P_n satisfies the hypothesis of Theorem 1.2. For $0 < \delta < 1$ let $\beta(\delta) = -\lg \delta$. Then

$$(2.4) \qquad \beta(\delta) \downarrow \qquad \text{as } \delta \uparrow ,$$

and

$$(2.5) \qquad \delta\beta(\delta) \uparrow \qquad \text{as } \delta \uparrow ,$$

provided $\delta \leq \frac{1}{4}$. Given an ε , with $0 < \varepsilon < A^{-1}$, let K_ε be the closed set

$$\{ x : \sup_{0<\delta\leq 1/4} \beta(\delta) M(x, \delta) \leq A/\varepsilon \} ,$$

where A is a positive constant to be determined later. Then K_ε is compact by Lemma 1.2, since $\beta^{-1}(\delta) \downarrow 0$ as $\delta \downarrow 0$. We will prove that $P_n(K_\varepsilon) \geq 1 - \varepsilon$ for all n by showing that

$$(2.6) \qquad P_n \{ x : \sup_{0<\delta\leq 1/4} \beta(\delta) M(x, \delta) \geq A/\varepsilon \} \leq \varepsilon .$$

If $x \in A_n$ it follows from (2.5) that

$$\sup_{0<\delta\le 2^{-n}} \beta(\delta)\, M(x,\delta) = \sup_{0<\delta\le 2^{-n}} \beta(\delta)\, \frac{\delta}{2^{-n}} \max_{0\le k<2^n} |x((k+1)2^{-n}) - x(k2^{-n})|$$

$$= \beta(2^{-n}) \max_{0\le k<2^n} |x((k+1)2^{-n}) - x(k2^{-n})| \ .$$

Hence, since $P_n(A_n) = 1$, since the differences (2.3) are normal and since $\varepsilon < A^{-1}$, we have

$$(2.7) \qquad P_n\{ \sup_{0<\delta\le 2^{-n}} \beta(\delta)\, M(x,\delta) \ge A/\varepsilon \}$$

$$\le \sum_{k=0}^{2^n-1} P_n\{ \beta(2^{-n})\, |x((k+1)2^{-n}) - x(k2^{-n})| \ge \frac{A}{\varepsilon} \}$$

$$\le \frac{\varepsilon}{A} \sqrt{\frac{2}{\pi}}\, \lg 2\; n\, 2^{n/2} \exp\left(- \frac{1}{2\lg^2 2} \cdot \frac{2^n}{n^2} \cdot \frac{A^2}{\varepsilon^2}\right)$$

$$\le \frac{\varepsilon}{A} A' ,$$

where

$$A' = \sup_{n\ge 1} \sqrt{\frac{2}{\pi}}\, \lg 2\; n\, 2^{n/2} \exp\left(- \frac{1}{2\lg^2 2} \cdot \frac{2^n}{n^2}\right) < \infty \ .$$

On the other hand from (2.4) and the fact that $M(x,\delta) \downarrow$ as $\delta \uparrow$, it follows that

$$(2.8) \qquad \sup_{2^{-n}<\delta\le 1/4} \beta(\delta) M(x,\delta) = \max_{2\le j\le n-2} \ \sup_{2^{j-1-n}<\delta\le 2^{j-n}} \beta(\delta) M(x,\delta)$$

$$\le \max_{2\le j\le n-2} \beta(2^{j-1-n}) M(x, 2^{j-n}) \ .$$

But if $x \in A_n$, a moment's reflection shows that

$$(2.9) \qquad M(x, 2^{j-n})$$

$$\leq 3 \max_{0 \leq i < 2^{n-j}} \sup \{ |x(t) - x(i2^{j-n})| : i2^{j-n} \leq t \leq (i+1)2^{j-n} \}$$

$$= 3 \max_{0 \leq i < 2^{n-j}} \max_{0 \leq k \leq 2^j} |x((i2^j + k)2^{-n}) - x(i2^{j-n})| .$$

Now by a well-known inequality concerning the maximum of the partial sums of independent symmetric random variables [8, p. 106] we have,

$$(2.10) \qquad P_n\{ 3\beta(2^{j-1-n}) \max_{0 \leq k \leq 2^j} |x((i2^j+k)2^{-n}) - x(i2^{j-n})| \geq A/\varepsilon \}$$

$$\leq 2 P_n\{ 3\beta(2^{j-1-n}) |x((i+1)2^{j-n}) - x(i2^{j-n})| \geq A/\varepsilon \}$$

$$\leq \frac{\varepsilon}{A} \frac{12 \lg 2}{\sqrt{2\pi}} \frac{n-j+1}{2^{(n-j)/2}} \exp\left(- \frac{1}{18 \lg^2 2} \frac{2^{n-j}}{(n-j+1)^2} \right) .$$

From (2.8), (2.9) and (2.10) it follows that

$$(2.11) \qquad P_n \{ \sup_{2^{-n} < \delta \leq 1/4} \beta(\delta) M(x, \delta) \geq A/\varepsilon \} \leq \frac{\varepsilon}{A} A'' ,$$

where

$$A'' = \frac{12 \lg 2}{\sqrt{2\pi}} \sum_{k=2}^{\infty} 2^{k/2} (k+1) \exp\left(- \frac{1}{18 \lg^2 2} \frac{2^k}{(k+1)^2}\right) < \infty .$$

If we take $A = A' + A''$ then (2.6) follows from (2.7) and (2.11).

We next show that

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

$$\int_C x_s x_t \, dP_n \to \min(s, t) \, . \tag{2.12}$$

Suppose that $s \leq t$ and let $j_n = [s2^n]$ and $k_n = [t2^n]$. If $x \in A_n$ then

$$x(s) = x(j_n 2^{-n}) + (s2^n - j_n)(x((j_n + 1)2^{-n}) - x(j_n 2^{-n})) \, ,$$

$$x(t) = x(k_n 2^{-n}) + (t2^n - k_n)(x((k_n + 1)2^{-n}) - x(k_n 2^{-n})) \, .$$

Hence, since the differences (2.3) are independent,

$$\int_C x_s x_t \, dP_n = \int_C x(j_n 2^{-n}) \, x(k_n 2^{-n}) \, dPn = j_n 2^{-n} \longrightarrow s \, .$$

Since $\{P_n\}$ satisfies the hypothesis of Theorem 1.2, there exists a sequence $\{n_\nu\}$ and a probability measure W on $\mathbf{C}$ such that $P_n \Longrightarrow W$. It follows easily from the normality of the variables (2.2) that W is Gaussian. And (2.1) follows from (2.12).

The method of proof of this theorem can be applied equally well in other cases. It is possible to use it to prove, for example, the existence of the stochastic process which arises in connection with Doob's approach to the Kolmogorov-Smirnov theorems [7].

We note at this point the well-known fact [8, p. 392] that

$$W\{x : \max_{a \leq t \leq b} (x(t) - x(a)) \geq \lambda \} = \sqrt{\frac{2}{\pi (b-a)}} \int_\lambda^\infty e^{-u^2/2(b-a)} \, du \, .$$

It follows from this that the W-measure of the boundary of any set of the form (1.29) is zero.

[illegible]

$$[illegible]$$

[illegible]

$$[illegible]$$

$$[illegible]$$

$$[illegible]$$

$$[illegible]$$

[illegible]
[illegible]
[illegible]
[illegible]
[illegible]
[illegible]
[illegible]
[illegible]
[illegible]

$$[illegible]$$

[illegible]
[illegible]

§3. A general invariance principle.

In this and the following sections we will be concerned with proving the convergence to W of the distributions (on C) of certain sequences of random polygons. Let $X_1, X_2, \ldots$ be a sequence of random variables on a probabaility space $(\Omega, \mathcal{B}, P)$. Let p_n be the random polygon defined by

$$p_n(t) = S_{j-1} + (nt-j+1)X_j \ , \ (j-1)n^{-1} \leq t \leq jn^{-1} \ , \quad j = 1, \ldots, n,$$

where $S_j = X_1 + \cdots + X_j$ and $S_0 = 0$. Thus p_n is the polygon with vertices at the points (jn^{-1}, S_j). As remarked at the end of §1, $P\{p_n \in A\}$ is, for $A \in \mathcal{C}$, a well defined quantity. Suppose there exists a sequence $\{a_n\}$ of positive constants such that if the measure P_n is defined by $P_n(A) = P\{a_n^{-1} p_n \in A\}$ then $P_n \Rightarrow W$. If this is true we say that the <u>invariance principle</u> holds for the sequence $\{X_n\}$ with <u>norming factors</u> a_n. It is the purpose of this section to derive a general set of conditions on $\{X_n\}$ under which the invariance principle holds. What the conditions lack in elegance they make up for in utility. In the subsequent sections we specialize these conditions in various ways.

For integers c, ν and n define

$$n_j = [jnc^{-1}] \ , \qquad j = 0, \ldots, c,$$

$$n_{j,u} = [n(\nu(j-1) + u)c^{-1}\nu^{-1}], \ j = 1, \ldots, c, \ u = 0, \ldots, \nu.$$

For any real numbers α_j, β_j with $\alpha_j \leq \beta_j$, $j = 1, \ldots, c$, let $E_{n,r}$ be the Ω set where the relations

$$(3.1) \qquad \alpha_j \leq a_n^{-1} S_i \leq \beta_j \qquad \text{if} \quad n_{j-1} < i \leq n_j$$

are satisfied for $i < r$, but not for $i = r$.

Theorem 3.1. The invariance principle holds for the sequence $\{X_n\}$ with norming factors a_n if the following two conditions are satisfied.

Condition (i). For each integer c the distribution of the random vector

$$(3.2) \qquad a_n^{-1} (S_{n_1}, S_{n_2} - S_{n_1}, \ldots, S_{n_c} - S_{n_{c-1}})$$

approaches, as $n \to \infty$, the normal distribution having zero means and having as covariance matrix c^{-1} times the cxc identity.

Condition (ii). For each integer c, each set $(\alpha_1, \ldots, \alpha_c, \beta_1, \ldots, \beta_c)$ and each $\varepsilon > 0$,

$$(3.3) \qquad \lim_{\nu \to \infty} \limsup_{n \to \infty} \sum_{r=1}^{n} P(E_{n,r} \cap \{ |S_{n_{j,u+1}} - S_r| \geq \varepsilon a_n \}) = 0,$$

where the $n_{j,u+1}$ corresponding to each r is defined by the relation

$$(3.4) \qquad n_{j,u} < r \leq n_{j,u+1} .$$

Proof. Throughout the rest of the paper we will be dealing with sums of the sort appearing in (3.3). In each instance $n_{j,u+1}$ is a function of r defined by (3.4).

To prove the theorem, let E_n be the Ω set where (3.1) is satisfied for all $i = 1, \dots, n$. Then $\Omega - E_n = \bigcup_{r=1}^{n} E_{n,r}$. Let E be the C set where

$$\alpha_j \leq x(t) \leq \beta_j$$

if $(j-1)c^{-1} \leq t \leq jc^{-1}$, for $j = 1, \dots, c$. Let D_ν be the C set where

$$\alpha_j \leq x(((j-1)\nu + u)c^{-1}\nu^{-1}) \leq \beta_j$$

for $j = 1, \dots, c$ and $u = 1, \dots, \nu$. Further, let F_n be the Ω set where

$$\alpha_j \leq a_n^{-1} S_{n_{j,u}} \leq \beta_j$$

for $j = 1, \dots, c$ and $u = 1, \dots, \nu$. Finally, let E_ε, $D_{\nu,\varepsilon}$ and $F_{n,\varepsilon}$ be defined in the same way as E, D_ν and F_n, but with α_j and β_j replaced by $\alpha_j + \varepsilon$ and $\beta_j - \varepsilon$ respectively.

For $n_{j,u} < r \leq n_{j,u+1}$ write

$$(3.5) \quad P(E_{n,r}) = P(E_{n,r} \cap \{|S_{n_{j,u+1}} - S_r| \geq \varepsilon a_n\})$$
$$+ P(E_{n,r} \cap \{|S_{n_{j,u+1}} - S_r| < a_n \varepsilon\}) .$$

Obviously the set in the second term of the right member of (3.5) is contained in $\Omega - F_{n,\varepsilon}$. Hence, since the $E_{n,r}$ are disjoint, we have

[illegible]

$$[illegible]$$

[illegible]

$$[illegible]$$

[illegible]

$$[illegible]$$

[illegible]

$$[illegible]$$

[illegible]

$$1 - P(E_n) = \sum_{r=1}^{n} P(E_{n,r}) \leq 1 - P(F_{n,\varepsilon}) + \zeta_{\nu, n} ,$$

where $\zeta_{\nu, n}$ is the sum in (3.3). Since $E_n \subset F_n$ we have

$$P(F_{n,\varepsilon}) - \zeta_{\nu, n} \leq P(E_n) \leq P(F_n) .$$

But

$$\lim_{n \to \infty} P(F_n) = W(D_\nu) ,$$

$$\lim_{n \to \infty} P(F_{n,\varepsilon}) = W(D_{\nu,\varepsilon}) ,$$

by Condition (i). Hence

$$W(D_{\nu,\varepsilon}) - \limsup_{n \to \infty} \zeta_{\nu, n} \leq \liminf_{n \to \infty} P_n(E_n) \leq \limsup_{n \to \infty} P_n(E_n) \leq W(D_\nu) .$$

Letting $\nu \to \infty$ we have, by Condition (ii),

$$W(E_\varepsilon) \leq \liminf_{n \to \infty} P(E_n) \leq \limsup_{n \to \infty} P(E_n) \leq W(E) .$$

Now $E_\varepsilon \uparrow E^o$ as $\varepsilon \downarrow 0$. Since $W(\tilde{E}) = 0$, $\lim_{n \to \infty} P(E_n) = W(E)$.

The proof of the invariance principle will be complete if we prove the following lemma, the hypothesis of which we have just shown to be satisfied.

Lemma 3.1. Let A_n be the set of polygons p which are linear on each of the intervals $((i - 1)n^{-1}, in^{-1})$ and satisfy $p(0) = 0$, and suppose P_n is a measure on $\mathcal{C}$ with $P(A_n) = 1$. Suppose further that

$$P_n(G_n) \longrightarrow W(E) \tag{3.6}$$

if E is any set of the form (1.29) and G_n is the set of $x \in C$ for which

$$\alpha_j \leq x(in^{-1}) \leq \beta_j$$

if $n_{j-1} < i \leq n_j$, for $i = 1, \ldots, n$. Then $P_n \Rightarrow W$.

Proof. Let ε be a small positive rational and let E_ε be the set where

$$\alpha_j \leq x(t) \leq \beta_j$$

if $(j-1)c^{-1} + \varepsilon \leq t \leq jc^{-1} - \varepsilon$ $(j = 2, \ldots, c)$, where

$$\alpha_1 \leq x(t) \leq \beta_1$$

$0 \leq t \leq c^{-1} - \varepsilon$, where

$$\alpha_c \leq x(t) \leq \beta_c$$

$1 - c^{-1} + \varepsilon \leq t \leq 1$, and where

$$\max(\alpha_j, \alpha_{j+1}) \leq x(t) \leq \min(\beta_j, \beta_{j+1})$$

if $jc^{-1} - \varepsilon \leq t \leq jc^{-1} + \varepsilon$ $(j = 1, \ldots, c-1)$. Analogously, define $G_{n,\varepsilon}$ to be the set where for $i = 1, \ldots, n$,

$$\alpha_j \leq x(in^{-1}) \leq \beta_j$$

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

if $(j - 1)c^{-1} + \varepsilon \leq in^{-1} \leq jc^{-1} - \varepsilon$, $(j = 2, \ldots, c - 1)$, where

$$\alpha_1 \leq x(in^{-1}) \leq \beta_1$$

if $0 \leq in^{-1} \leq c^{-1} - \varepsilon$, where

$$\alpha_c \leq x(in^{-1}) \leq \beta_c$$

if $1 - c^{-1} + \varepsilon \leq in^{-1} \leq 1$, and where

$$\max(\alpha_j, \alpha_{j+1}) \leq x(in^{-1}) \leq \min(\beta_j, \beta_{j+1})$$

if $jc^{-1} - \varepsilon \leq in^{-1} \leq jc^{-1} + \varepsilon$ $(j = 1, \ldots, c - 1)$. Since ε is rational, E_ε can be cast in the form (1.29) and $G_{n,\varepsilon}$ bears the same relation to E_ε as G_n does to E. Then by hypothesis

$$P_n(G_{n,\varepsilon}) \longrightarrow W(E_\varepsilon) \tag{3.7}$$

as $n \longrightarrow \infty$. Now $E \subset G_n$, while $G_{n,\varepsilon} A_n \subset E$ provided $n^{-1} < \varepsilon$. Hence

$$P_n(G_{n,\varepsilon}) \leq P_n(E) \leq P_n(G_n)$$

for large n and, by (3.6) and (3.7),

$$W(E_\varepsilon) \leq \liminf_{n \to \infty} P_n(E) \leq \limsup_{n \to \infty} P_n(E) \leq W(E). \tag{3.8}$$

Now $E_\varepsilon \uparrow F$ as $\varepsilon \downarrow 0$, where $W(E - F) = 0$. Letting $\varepsilon \longrightarrow 0$ in (3.8) we have $\lim_{n \to \infty} P_n(E) = W(E)$. Hence $P_n \Rightarrow W$ by Theorem 1.3. This completes the proof of the lemma, and hence of the theorem.

With only slight complications Lemma 3.1 can be proved with W replaced by an arbitrary limiting measure.

§4. The invariance principle for Markov processes.

In this section we prove, using Theorem 3.1, the invariance principle for discrete Markov processes satisfying Doeblin's condition. We use the definitions, notations and results of [8, Ch. V].

Let X be a space of points ξ and let $\mathcal{F}_X$ be a Borel field of subsets of X. Let $\{x_n, n \geq 1\}$ be a Markov process with state space X and stationary transition probabilities

$$p(\xi, A) = P\{x_{n+1} \in A \,||\, x_n = \xi\}. \tag{4.1}$$

That is, $\{x_n\}$ is a sequence of measurable functions from some probability space $(\Omega, \mathcal{B}, P)$ to X, such that (4.1) holds, where the transition function p is a measurable function of ξ for a fixed $A \in \mathcal{F}_X$ and is a probability measure on $\mathcal{F}_X$ for fixed ξ. The initial distribution π is defined by

$$\pi(A) = P\{x_1 \in A\},$$

and the n-step transition probabilities by

$$p^{(n)}(\xi, A) = P\{x_{n+1} \in A \,||\, x_1 = \xi\}.$$

The existence problems involved here are resolved in [8].

We assume that the process satisfies the hypothesis of Doeblin:

[illegible]

[illegible]

$$[illegible]$$

[illegible]

[illegible]

$$[illegible]$$

[illegible]

Hypothesis (D). There exists a finite-valued measure φ on $\mathcal{F}_X$, an integer $\nu \geq 1$ and a positive ε, such that if $\varphi(A) \leq \varepsilon$ then

$$p^{(\nu)}(\xi, A) \leq 1 - \varepsilon$$

for all $\xi \in X$.

Note that (D) is a hypothesis on the transition function alone and is independent of the initial distribution π.

It is shown in [8] how, under (D), the states ξ can be classified according to their ergodic properties. It is shown further that the ergodic theorem holds if the following hypothesis is satisfied.

Hypothesis (D_o).

(a) Hypothesis (D) is satisfied.

(b) There is only a single ergodic set and this contains no cyclically moving subsets.

That is, it is shown that if (D_o) holds then there exist positive constants r and ρ, $\rho < 1$, and a (unique) stationary initial distribution p such that

$$|p^{(n)}(\xi, E) - p(E)| \leq r\rho^n$$

for all $\xi \in X$, $E \in \mathcal{F}_X$ and $n \geq 1$. The results of this section will be obtained under the assumption of (D_o).

In what follows the initial distribution under the assumption of which a probability is computed will be denoted by a subscript, thus: $P_\pi(E)$. If $\pi = p$, the stationary initial distribution, the subscript will be omitted. In statements involving only transition probabilities, e.g., (4.1), the initial distribution is irrelevant and the subscript will be omitted in any case.

We state for reference three results proved in [8, p. 224].

<u>Lemma 4.1.</u> Under Hypothesis (D_0), if f is a bounded (perhaps complex-valued) random variable, $|f| \leq M$, on $x_{k+1}, x_{k+2}, \ldots$ sample space, then

$$(4.2) \qquad |E\{f \mid\mid x_1\} - \{f\}| \leq 2\,r\,M\rho^k .$$

In several of the subsequent applications of this lemma f will be the characteristic function of a set.

<u>Lemma 4.2.</u> Under Hypothesis (D_0), let f be real-valued function of ξ, measurable $\mathcal{F}_X$, with

$$E\{f(x_1)\} = 0, \quad E\{(f(x_1))^2\} = \sigma^2 < \infty .$$

Then as $n \to \infty$,

$$E\{(\sum_{j=1}^{n} f(x_j))^2\} \sim n\,\sigma_1^2 ,$$

where σ_1^2 is a constant depending on f and on the process.

Lemma 4.3. Under Hypothesis (D_o), let f be a real-valued function of ξ, measurable $\mathcal{F}_X$, with

$$E\{f(x_1)\} = 0, \quad E\{|f(x_1)|^{2+\delta}\} < \infty$$

for some $\delta \geq 0$. Then there is a constant a, for which

$$E\{|\sum_{j=1}^{n} f(x_j)|^{2+\delta}\} \leq a\, n^{1+(\delta/2)}, \qquad n = 1, 2, \ldots \quad .$$

It is convenient to have available the following corollary of Lemma 4.1. Suppose we have positive integers u_i, v_i with

$$(4.3) \qquad u_1 \leq v_1 < u_2 \leq v_2 < \ldots < u_m \leq v_m \quad .$$

Suppose further that

$$(4.4) \qquad u_i - v_{i-1} \geq B \geq 1, \qquad i = 2, \ldots, m \quad .$$

Lemma 4.4. Under Hypothesis (D_o), let f_j be a (perhaps complex-valued) random variable, with $|f_j| \leq 1$, on $x_{u_j}, \ldots, x_{v_j}$ sample space, for $j = 1, \ldots, m$. If (4.3) and (4.4) hold, then

$$(4.5) \qquad |E\{f_1 \cdots f_m\} - E\{f_1\} \cdots E\{f_m\}| \leq 2m\gamma\rho^B \quad .$$

Proof. The proof goes by induction on m. The result being trivial for $m = 1$, assume it is true for some $m - 1$. Then

$$E\{f_1 \cdots f_m\} = E\{f_1 \cdots f_{m-1}\} E\{f_m \| x_{v_{m-1}}\} = E\{f_1 \cdots f_{m-1}\} E\{f_m\} + \varepsilon,$$

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

where $|\varepsilon| \leq 2\, r \rho^{B}$ by Lemma 4.1 and the bound $|f_j| \leq 1$. And now (4.5) follows from the induction hypothesis.

We come now to the invariance principle.

Theorem 4.1. Under Hypothesis (D_0), let f be a real-valued function of ξ, measurable $\mathcal{F}_X$, with

$$E\{f(x_1)\} = 0, \quad E\{|f(x_1)|^{2+\delta}\} < \infty$$

for some $\delta > 0$. Then

$$(4.6) \qquad \lim_{n\to\infty} E\{(n^{-1/2} \sum_{j=1}^{n} f(x_j))^2\} = \sigma_1^2$$

exists. If $\sigma_1^2 > 0$ then the invariance principle holds for the sequence $\{f(x_n)\}$ with norming factors $\sigma_1 n^{1/2}$, no matter what the initial distribution π.

Proof. That the limit (4.6) exists is simply a restatement of Lemma 4.2. We prove the result first under the assumption of stationarity and remove this restriction later.

We must show that Conditions (i) and (ii) of Theorem 3.1 are satisfied. In the notation of that theorem, we must first prove that the distribution of the vector

$$\sigma_1^{-1} n^{-1/2} (S_{n_1}, S_{n_2} - S_{n_1}, \ldots, S_{n_c} - S_{n_{c-1}})$$

approaches the appropriate normal distribution, where $S_k = f(x_1) + \cdots + f(x_k)$. Our proof of this part of the theorem follows [8]. Let $\{\alpha_n\}$ and $\{\beta_n\}$ be two sequences of positive

integers such that if $\{\mu_n\}$ is defined by

$$\mu_n = [\,(\min_{1 \le j \le c} (n_j - n_{j-1}) - \beta_n)(\alpha_n + \beta_n)^{-1}\,] , \tag{4.7}$$

then

$$\lim_{n \to \infty} \mu_n \beta_n \alpha_n^{-1} = 0 , \tag{4.8}$$

$$\lim_{n \to \infty} \mu_n \rho^{\beta_n} = 0 , \tag{4.9}$$

while α_n, β_n and μ_n all go to infinity. For example one can take $\beta_n^4 \sim nc^{-1}$ and $\alpha_n = \beta_n^3$.

Now for $j = 1, \ldots, c$ let

$$y_{j,m} = \sum_{i=(m-1)(\alpha+\beta)+1}^{(m-1)(\alpha+\beta)+\alpha} f(x_{n_{j-1}+i}) , \qquad m = 1, \ldots, \mu ,$$

$$y'_{j,m} = \sum_{i=(m-1)(\alpha+\beta)+\alpha+1}^{m(\alpha+\beta)} f(x_{n_{j-1}+i}) , \qquad m = 1, \ldots, \mu ,$$

$$y'_{j,\mu+1} = \sum_{i=\mu(\alpha+\beta)+1}^{n_j} f(x_{n_{j-1}+i}) .$$

We prove that

$$\operatorname*{p\,lim}_{n \to \infty} n^{-1/2} \sum_{m=1}^{\mu+1} y'_{j,m} = 0 , \qquad j = 1, \ldots, c . \tag{4.10}$$

Now by Lemma 4.2, Minkowski's inequality and (4.8),

$$E^{1/2}\{(n^{-1/2}\sum_{m=1}^{\nu+1} y'_{j,m})^2\} < n^{-1/2}\sum_{m=1}^{\nu+1} E^{1/2}\{(y'_{j,m})^2\}$$

$$\leq An^{-1/2}(\nu\beta^{1/2} + (\alpha+2\beta)^{1/2}) \leq A\nu^{-1/2}\alpha^{-1/2}(\nu\beta^{1/2} + (\alpha+2\beta)^{1/2}) \to 0,$$

where A is a constant. This implies (4.10).

Hence by Theorem A.4 it suffices to prove the asymptotic normality of the random vector

$$\sigma_1^{-1}n^{-1/2}\left(\sum_{m=1}^{\nu} y_{1,m}, \ldots, \sum_{m=1}^{\nu} y_{c,m}\right). \tag{4.11}$$

Let

$$\varphi_n(u_1,\ldots,u_c) = E\{\exp(i\sum_{j=1}^{c} u_j \sum_{m=1}^{\nu} \sigma_1^{-1} n^{-1/2} y_{j,m})\}$$

be the characteristic function of (4.11). Now the last term $f(x_i)$ occurring in $y_{j,m}$ and the first occurring in $y_{j,m+1}$ have β such terms in between them. And the last term of $y_{j,\nu}$ and the first of $y_{j+1,1}$ have at least β others in between. Hence by Lemma 4.4,

$$\varphi_n(u_1,\ldots,u_c) = \prod_{m=1}^{\nu}\prod_{j=1}^{c} E\{\exp(it_j y_{j,m})\} + \varepsilon_n, \tag{4.12}$$

where $|\varepsilon_n| \leq 2\gamma c\nu\rho^{\beta+1} \to 0$ by (4.9). Let

$z_{j,m}$ $(j = 1, \ldots, c,\ m = 1, \ldots, \mu)$ be independent random variables each having the distribution of $y_{1,1}$. By (4.12), the proof of Condition (i) will be complete if we show that the distribution of the vector

$$\sigma_1^{-1} n^{-1/2} \left(\sum_{m=1}^{\mu} z_{1,m}, \ldots, \sum_{m=1}^{\mu} z_{c,m} \right) \tag{4.13}$$

approaches the appropriate normal distribution. Since $\mu / \alpha n \to c^{-1}$ by (4.8), it follows that

$$\lim_{n \to \infty} E\{(\sigma_1^{-1} n^{-1/2} \sum_{m=1}^{\mu} z_{j,m})^2\} = c^{-1}, \qquad j = 1, \ldots, c . \tag{4.14}$$

Since the components of (4.13) are independent, its covariance matrix approaches c^{-1} times the $c \times c$ identity matrix. By Theorem A.2 it suffices to show that Lyapounov's condition is satisfied. By Lemma 4.3 there is a constant a such that

$$E\{|z_{j,m}|^{2+\delta}\} \leq a\, \alpha^{1+(\delta/2)} .$$

Hence, by (4.14), for n sufficiently large,

$$\frac{\sum_{m=1}^{\mu} E\{|\sigma_1^{-1} n^{-1/2} z_{j,m}|^{2+\delta}\}}{E\{(\sigma_1^{-1} n^{-1/2} \sum_{m=1}^{\mu} z_{j,m})^2\})^{1+(\delta/2)}} \leq 2\mu^{-\delta/2} \to 0 .$$

Thus Lyapounov's condition holds.

Having shown that Condition (i) of Theorem 3.1 is satisfied,

we turn to Condition (ii). Let $E_{n,r}$, $n_{j,u+1}$, c and ν be as in that theorem. Define a sequence $\{\beta_n\}$ of integers by $\beta_n = [\lg n]$. If $r+\beta < n_{j,u+1}$ then

$$\text{(4.15)} \qquad \begin{aligned} &P(E_{n,r} \cap \{|S_{n_{j,u+1}} - S_r| \geq \varepsilon n^{1/2}\}) \\ &\leq P(E_{n,r} \cap \{|S_{n_{j,u+1}} - S_{r+\beta}| \geq \varepsilon n^{1/2}/2\}) \\ &\quad + P\{|S_{r+\beta} - S_r| \geq \varepsilon n^{1/2}/2\}, \end{aligned}$$

and we can estimate the terms on the right separately. Now

$$\text{(4.16)} \qquad \begin{aligned} &P\{|S_{n_{j,u+1}} - S_{r+\beta}| \geq \varepsilon n^{1/2}/2 \mid\mid x_1, \ldots, x_r\} \\ &\leq P\{|S_{n_{j,u+1}} - S_{r+\beta}| \geq \varepsilon n^{1/2}/2\} + 2r\rho^{\beta} \end{aligned}$$

by Lemma 4.1 and the Markov property. And by Chebyshev's inequality and Lemma 4.2,

$$\text{(4.17)} \qquad \begin{aligned} &P\{|S_{n_{j,u+1}} - S_{r+\beta}| \geq \varepsilon n^{1/2}/2\} \\ &\leq A\varepsilon^{-2} n^{-1}(n_{j,u+1} - (r+\beta)) \leq 2A/\varepsilon^2 c\,\nu, \end{aligned}$$

where A is a constant. By (4.16), (4.17) and the defining property of conditional probabilities,

[illegible]

[illegible]

[illegible]

$$[illegible] \quad (\text{[illegible]})$$

$$[illegible]$$

$$[illegible]$$

[illegible]

$$[illegible] \quad (\text{[illegible]})$$

[illegible]

[illegible]

[illegible]

$$[illegible] \quad (\text{[illegible]})$$

$$[illegible]$$

[illegible]

[illegible]

$$(4.18)\qquad P(E_{n,r} \cap \{|S_{n_{j,u+1}} - S_{r+\beta}| \geq \varepsilon\, n^{1/2}/2\})$$

$$\leq \left(\frac{2A}{\varepsilon^2 c\,\nu} + 2\,\gamma\rho^{\beta}\right) P(E_{n,r}) .$$

To estimate the second term in (4.15) observe that

$$(4.19)\qquad P\{|S_{r+\beta} - S_r| \geq \varepsilon\, n^{1/2}/2\}$$

$$\leq \sum_{i=r+1}^{r+\beta} P\{|f(x_i)| \geq \varepsilon\, \beta^{-1} n^{1/2}/2\}$$

$$= \beta P\{|f(x_1)| \geq \varepsilon\, \beta^{-1} n^{1/2}/2\} .$$

Therefore, by (4.15), (4.18) and (4.19),

$$P(E_{n,r} \cap \{|S_{n_{j,u+1}} - S_r| \geq \varepsilon\, n^{1/2}\})$$

$$\leq \left(\frac{2A}{\varepsilon^2 c\,\nu} + 2\,\gamma\,\rho^{\beta}\right) P(E_{n,r}) + \beta P\{|f(x_1)| \geq \varepsilon\, \beta^{-1} n^{1/2}/2\} .$$

This estimate was obtained under the assumption that $r + \beta < n_{j,u+1}$, but obviously holds in the other case as well. Since the $E_{n,r}$ are disjoint we have then,

$$(4.20)\qquad \sum_{r=1}^{n} P(E_{n,r} \cap \{|S_{n_{j,u+1}} - S_r| \geq \varepsilon\, n^{1/2}\})$$

$$\leq \frac{2A}{\varepsilon^2 c\,\nu} + 2\,\gamma\,\rho^{\beta} + n\beta P\{|f(x_1)| \geq \varepsilon\, \beta^{-1} n^{1/2}/2\} .$$

Now

$$n\beta P\{|f(x_1)| \geq \varepsilon \beta^{-1} n^{1/2}/2\}$$

$$\leq \left(\frac{2}{\varepsilon}\right)^{2+\delta} \beta^{3+\delta} n^{-\delta/2} E\{|f(x_1)|^{2+\delta}\} .$$

Hence as $n \to \infty$ the second and third terms on the right in (4.20) go to zero, and (3.3) follows immediately.

We have thus proved the theorem under the assumption that the initial distribution is the stationary one, which assumption we now remove. Let π be any initial distribution. We show first of all that there exists a sequence $\{\beta_n\}$ of integers going to infinity so slowly that

$$\text{(4.21)} \qquad \lim_{n \to \infty} P_\pi \{ \max_{i \leq \beta_n} |S_i| \geq \varepsilon n^{1/2} \} = 0$$

for all $\varepsilon > 0$. For each k select an integer m_k so that

$$\sum_{i=1}^{k} P_\pi \{|f(x_i)| \geq k^{-1} n^{1/3}\} < k^{-1} .$$

if $n \geq m_k$. Clearly we can choose the m_k so that $m_k < m_{k+1}$. And now let $\beta_n = k$ if $m_k < n \leq m_{k+1}$. Then β_n goes to infinity and if $m_k < n \leq m_{k+1}$ then

$$P_\pi\{ \max_{i \leq \beta_n} |S_i| \geq n^{1/3}\} \leq \sum_{i=1}^{k} P_\pi \{|f(x_i)| \geq k^{-1} n^{1/3}\}$$

$$< k^{-1} = \beta_n^{-1} .$$

[illegible]

$$[illegible]$$

[illegible]

$$[illegible]$$

[illegible]

$$[illegible]$$

[illegible]

But this implies (4.21) . It is obvious that we can also choose β so that

$$\beta = o(n^{1/3}) \quad . \tag{4.22}$$

Let p_n be the polygon defined by

$$p_n(t) = S_{j-1}(j-nt) + S_j(nt-j+1), \quad (j-1)n^{-1} \leq t \leq jn^{-1}, \quad j = 1, \ldots, n,$$

where $S_0 = 0$. And let p_n' be the polynomial defined by

$$p_n'(t) = \begin{cases} p_n(t) & \text{if} \quad 0 \leq t \leq \beta_n n^{-1} \\ p_n(\beta_n n^{-1}) & \text{if} \quad \beta_n n^{-1} \leq t \leq 1 , \end{cases}$$

where $\{\beta_n\}$ satisfies (4.21) and (4.22) . Finally, let $p_n'' = p_n - p_n'$. Now (4.21) implies that

$$\lim_{n \to \infty} P_\pi \{ \max_{0 \leq t \leq 1} n^{-1/2} p_n'(t) \geq \varepsilon \} = 0 \tag{4.23}$$

for all $\varepsilon > 0$. Also

$$P\{ \max_{i \leq \beta} |S_i| \geq \varepsilon n^{1/2} \} \leq \beta P\{ |f(x_1)| \geq \beta^{-1} \varepsilon n^{1/2} \}$$

$$\leq \frac{\beta^3}{\varepsilon^2 n} E\{(f(x_1))^2\} \quad ,$$

so that by (4.22)

(4.24) $$\lim_{n\to\infty} P\{ \max_{0\le t\le 1} n^{-1/2} p_n'(t) \geq \varepsilon \} = 0$$

for all $\varepsilon > 0$. Since we have shown that the invariance principle holds if π is the stationary distribution,

$$P\{ \sigma_1^{-1} n^{-1/2} p_n \in A \} \longrightarrow W(A)$$

for all $A \in \mathcal{C}$ for which $W(\tilde{A}) = 0$. By (4.24) and Theorem 1.4,

(4.25) $$P\{ \sigma_1^{-1} n^{-1/2} p_n'' \in A \} \longrightarrow W(A) .$$

It is clear that the set $\{\sigma_1^{-1} n^{-1/2} p_n'' \in A\}$ is measurable on $x_{\beta+1}, x_{\beta+2}, \ldots$ sample space. Hence by Lemma 4.1,

$$|P_\pi \{\sigma_1^{-1} n^{-1/2} p_n'' \in A\} - P\{\sigma_1^{-1} n^{-1/2} p_n'' \in A\}|$$

$$\leq 2\gamma\rho^\beta \longrightarrow 0 .$$

By (4.25) then,

$$P_\pi \{\sigma_1^{-1} n^{-1/2} p_n'' \in A\} \longrightarrow W(A) .$$

But from this, (4.23) and Theorem 1.4 it follows that if $W(\tilde{A}) = 0$ then

$$P_\pi \{\sigma_1^{-1} n^{-1\ 2} p_n \in A\} \longrightarrow W(A) .$$

We have thus proved the invariance principle with no restrictions on π.

§5. The invariance principle for m-dependent random variables.

A sequence $\{X_n\}$ of random variables is said to be m-dependent if the random vectors $(X_n, \ldots, X_{n+r})$ and $(X_{n+s}, \ldots, X_{n+t})$ are independent whenever $s - r > m$. Sequences having this property are of interest in statistics and have been studied by various authors (cf. Bernstein [2], Hoeffding and Robbins [14], Diananda [5] and Marsaglia [18].) In this section we prove that the invariance principle holds for m-dependent sequences if one or the other of two auxilliary conditions is satisfied.

Then let $\{X_n\}$ be an m-dependent sequence of random variables with zero means and finite variances. Let $S_n = X_1 + \ldots + X_n$ and $s_n^2 = E\{S_n^2\}$.

Theorem 5.1. If, for an m-dependent sequence $\{X_n\}$, $E\{X_n^2\}$ is bounded,

$$(5.1) \qquad |s_n^2 - n\sigma^2| = 0(1)$$

for some constant $\sigma^2 > 0$, and

$$(5.2) \qquad \lim_{n \to \infty} s_n^{-(2+\delta)} \sum_{i=1}^{n} E\{|X_i|^{2+\delta}\} = 0$$

for some $\delta > 0$, then the invariance principle holds for the sequence $\{X_n\}$ with norming factors $\sigma n^{1/2}$.

Proof. We first show that Condition (i) of Theorem 3.1 is satisfied, using the technique of Marsaglia [18] (cf. Theorem A.4 below). Let $(n_1, \ldots, n_c)$ be defined as in §3, and for each pair (n, k) with $2m < k < nc^{-1}$ and each $j = 1, \ldots, c$ define

$$y_{j,i} = \sum_{v=1}^{k-m} X_{n_{j-1}+(i-1)k+v}, \qquad 1 \leq i \leq [k^{-1}(n_j - n_{j-1})],$$

$$y'_{j,i} = \sum_{v=1}^{m} X_{n_{j-1}+ik-m+v}, \qquad 1 \leq i < [k^{-1}(n_j - n_{j-1})],$$

$$y'_{j,i} = \sum_{v=1}^{n_j-n_{j-1}-ik+m} X_{n_{j-1}+ik-m+v}, \qquad i = [k^{-1}(n_j - n_{j-1})].$$

Let

$$g_{n,k}^{(j)} = \sum_{i=1}^{[k^{-1}(n_j-n_{j-1})]} y_{j,i}, \qquad e_{n,k}^{(j)} = \sum_{i=1}^{[k^{-1}(n_j-n_{j-1})]} y'_{j,i}.$$

Now by Hölder's inequality, if $i < [k^{-1}(n_j - n_{j-1})]$ then

$$E\{(y'_{j,i})^2\} \leq m \sum_{v=1}^{m} E\{X^2_{n_{j-1}+ik-m+v}\} \leq m^2 B,$$

where B is the bound on $E\{X_n^2\}$. Using this inequality, a similar one for the case $i = [k^{-1}(n_j - n_{j-1})]$ and the fact that the $y'_{j,i}$ are independent, we see that

$$(5.3) \qquad E\{(e_{n,k}^{(j)})^2\} \leq (n_j - n_{j-1})k^{-1} m^2 B + (k+m)^2 B.$$

One obtains in a similar manner the inequality

(5.4) $$|E\{(S_{n_j} - S_{n_{j-1}})^2\} - E\{(g^{(j)}_{n,k})^2\}|$$
$$\leq 5(n_j - n_{j-1})k^{-1}m^2 B + (k+m)^2 B + 2m^2 B \ .$$

From (5.3) and (5.4) it follows that

(5.5) $$\lim_{k \to \infty} \ \limsup_{n \to \infty} \ n^{-1} \sigma^{-2} E\{(e^{(j)}_{n,k})^2\} = 0$$

(5.6) $$\lim_{k \to \infty} \ \limsup_{n \to \infty} \ n^{-1} \sigma^{-2} |E\{(S_{n_j} - S_{n_{j-1}})^2\} - E\{(g^{(j)}_{n,k})^2\}| = 0 \ .$$

Note that these two relations have been obtained without the use of (5.2).

Now by (5.1) ,

(5.7) $$E\{(S_{n+i} - S_n)^2\} = i\,\sigma^2 + \vartheta \ ,$$

where ϑ is bounded. From this fact and (5.6) it follows that

(5.8) $$\lim_{k \to \infty} \ \lim_{n \to \infty} c\,n^{-1} \sigma^{-2} E\{(g^{(j)}_{n,k})^2\} = 1 \ ,$$

where the iterated limit is to be taken in the strong sense (cf. the appendix).

We now show that if k is sufficiently large than the distribution of the vector

$$((\tau^{(1)}_{n,k})^{-2} \ g^{(1)}_{n,k} \ , \ \ldots , \ (\tau^{(c)}_{n,k})^{-2} \ g^{(c)}_{n,k}) \ ,$$

where $(\tau^{(j)}_{n,k})^2 = E\{(g^{(j)}_{n,k})^2\}$, approaches, as $n \to \infty$, the

normal distribution having zero means and having the $c \times c$ identity as its covariance matrix. By Theorem A.2 it suffices to show that Lyapounov's condition is satisfied. But for k sufficiently large this follows from (5.7), (5.1) and the fact that

$$E\{|y_{i,i}|^{2+\delta}\} \leq k^{\delta} \sum_{v=1}^{k-m} E\{|X_{n_{j-1}+(i-1)k+v}|^{2+\delta}\} .$$

But now from (5.8) and Theorem A.5 it follows that the distribution of

$$\sigma^{-1} n^{-1/2} (g_{n,k}^{(1)}, \dots, g_{n,k}^{(c)})$$

approaches, if $n \to \infty$ and then $k \to \infty$, the normal distribution having zero means and having c^{-1} times the identity as covariance matrix. In order to show that the distribution of

$$\sigma^{-1} n^{-1/2} (S_{n_1}, S_{n_2} - S_{n_1}, \dots, S_{n_c} - S_{n_{c-1}})$$

approaches, as $n \to \infty$, this same normal distribution it is enough, by Theorem A.4, to show that

$$\text{(5.10)} \qquad \underset{k \to \infty}{\text{p lim}} \; \underset{n \to \infty}{\text{p lim}} \; \sigma^{-1} n^{-1/2} (e_{n,k}^{(1)}, \dots, e_{n,k}^{(c)}) = 0 ,$$

since $S_{n_j} - S_{n_{j-1}} = g_{n,k}^{(j)} + e_{n,k}^{(i)}$. But (5.10) follows immediately from (5.5) and Chebyshev's inequality.

We have thus proved Condition (i) of Theorem 3.1, and pass on to Condition (ii). Using all the notations of that theorem,

$$P\{|S_{r+m} - S_r| \geq \varepsilon\, n^{1/2}/2\} \leq \sum_{\nu=r}^{r+m} P\{|X_\nu| \geq \varepsilon\, n^{1/2}/2m\}\,.$$

Hence,

(5.11) $$\sum_{r=1}^{n} P\{|S_{r+m} - S_r| \geq \varepsilon\, n^{1/2}/2\} \leq m \sum_{r=1}^{n} P\{|X_\nu| \geq \varepsilon\, n^{1/2}/2m\}$$

$$\leq m\left(\frac{2m}{\varepsilon}\right)^{2+\delta} \frac{1}{n^{1+(\delta/2)}} \sum_{r=1}^{n} E\{|X_r|^{2+\delta}\} \longrightarrow 0$$

as $n \longrightarrow \infty$, by (5.2). And (making the inessential assumption that $r+m < n_{j,u+1}$) by Chebyshev's inequality and (5.7),

$$P\{|S_{n_{j,u+1}} - S_{r+m}| \geq \varepsilon\, n^{1/2}/2\} \leq \frac{4}{\varepsilon^2 c \nu} + \frac{4}{\varepsilon^2 n}\, \vartheta\,.$$

Hence, since $S_{n_{j,u+1}} - S_{r+m}$ is independent of $E_{n,r}$,

(5.12) $$\limsup_{n \longrightarrow \infty} \sum_{r=1}^{n} P(E_{n,r} \cap \{|S_{n_{j,u+1}} - S_{r+m}| \geq \varepsilon\, n^{1/2}/2\})$$

$$\leq 4/c\nu\varepsilon^2.$$

And now (3.3) follows from (5.11), (5.12) and

(5.13) $$P(E_{n,r} \cap \{|S_{n_{j,u+1}} - S_r| \geq \varepsilon\, n^{1/2}\})$$

$$\leq P\{|S_{r+m} - S_r| \geq \varepsilon\, n^{1/2}/2\} + P(E_{n,r} \cap \{|S_{n_{j,u+1}} - S_{m+r}|$$

$$\geq \varepsilon\, n^{1/2}/2\})\,.$$

It is possible, at the expense of complicating somewhat the proof of Condition (ii), to relax the condition (5.1). In particular, it can be replaced by $s_n^2 \sim n\,\sigma^2$.

Theorem 5.2. If $\{X_n\}$ is a stationary m-dependent sequence of random variables with zero means and finite variances, then the invariance principle holds for $\{X_n\}$ with norming factors $n^{1/2}\sigma$, where

$$\sigma^2 = E\{X_1^2\} + 2 \sum_{k=1}^{m} E\{X_1 X_{k+1}\} .$$

Proof. Define $y_{j,i}$, $y'_{j,i}$, $g_{n,k}^{(j)}$, and $e_{n,k}^{(j)}$ as in the proof of Theorem 5.1. It is a simple matter to show that (5.1) holds here. Since $E\{X_n^2\}$ is bounded it follows that (5.5) and (5.6) hold in this case as well. In order to establish that Condition (i) of Theorem 3.1 holds in the present case it suffices to show that the vector (5.4) is asymptotically normal. But this follows immediately by Theorem A.3.

To prove the Condition (ii) holds we proceed as before. In fact (5.13) and (5.12) are still valid. Finally,

$$\sum_{r=1}^{n} P\{|S_{r+m} - S_r| \geq \varepsilon\, n^{1/2}/2\} \leq mn\,P\{|X_1| \geq \varepsilon\, n^{1/2}/2m\} ,$$

and the right-hand side of this inequality goes to zero since $E\{X_1^2\} < \infty$.

An immediate consequence of this theorem is the original result of Donsker [6].

<u>Theorem 5.3.</u> If $\{X_n\}$ is an independent sequence of random variables which are identically distributed with zero mean and finite variance σ^2, then the invariance principle holds for $\{X_n\}$ with norming factors $n^{1/2}\sigma$.

§6. The invariance principle for linear processes with m-dependent residuals.

Let $\{Y_j; -\infty < j < \infty\}$ be an m-dependent process and $\{A_t, t \geq 0\}$ a sequence of constants such that for every integer i, $\sum_{t=0}^{n} A_t Y_{i-t}$ converges in probability to a variable $X_i = \Sigma_{t=0}^{\infty} A_t Y_{i-t}$ as $n \to \infty$. Then we say that $\{X_i, -\infty < i < \infty\}$ is a discrete linear process with m-dependent residuals. Processes of this sort are of interest in the analysis of time series (cf. Diananda [5] for references to the statistical literature).

Suppose that $\{Y_j, -\infty < j < \infty\}$ is a stationary m-dependent process such that Y_j has zero mean and finite variance and suppose that

(6.1) $$\Sigma_{t=0}^{\infty} |A_t| < \infty .$$

For a fixed i, let $T_n = \Sigma_{t=0}^{n} A_t Y_{i-t}$. It is easy to show, using the m-dependence and stationarity properties of $\{Y_j\}$, that

$$E\{(T_{n+u} - T_n)^2\} \leq E\{Y_1^2\} \left(\sum_{t=n}^{\infty} A_t^2 + 2mA \sum_{t=n}^{\infty} |A_t| \right).$$

[illegible]

Therefore

$$\lim_{n\to\infty} \sup_{u>0} E\{(T_{n+u} - T_n)^2\} = 0 ,$$

and $\{T_n\}$ is fundamental, and hence convergent, in probability. Thus $\Sigma_{t=0}^{n} A_t Y_{i-t}$ converges in probability, as $n \to \infty$, to some random variable $X_i = \Sigma_{t=0}^{\infty} A_t Y_{i-t}$, and $\{X_i, -\infty < i < \infty\}$ is a discrete linear process with m-dependent residuals. It is trivial to show that $\{X_i\}$ is stationary. In [5] Diananda has shown that the central limit theorem holds for processes $\{X_i\}$ which arise in this way, that is, if $\{Y_j\}$ is m-dependent and stationary, Y_j has finite variance and (6.1) holds. It is the purpose of this section to prove the invariance principle for such processes. We are forced, however, to make a stronger assumption on the nature of the sequence $\{A_t\}$, viz., we assume that

$$(6.2) \qquad |A_n| = 0(n^{-3}) .$$

At the end of the proof we indicate some ways in which the requirement (6.2) can be relaxed.

<u>Theorem 6.1.</u> Let $\{Y_j, -\infty < j < \infty\}$ be a stationary m-dependent process with zero means and finite variances and assume

that (6.2) holds. Then $\Sigma_{t=0}^{n} A_t Y_{i-t}$ converges, as $n \to \infty$, in probability to some random variable $X_i = \Sigma_{t=0}^{\infty} A_t Y_{i-t}$, so that X_i is a stationary discrete linear process with m-dependent residuals, and the invariance principle holds for $\{X_i\}$ with norming factors $n^{1/2}\sigma$, where

$$(6.3) \qquad \sigma^2 = \left(\sum_{i=0}^{\infty} A_i\right)^2 \left(E\{Y_0^2\} + 2\sum_{v=1}^{m} E\{Y_0 Y_v\}\right).$$

<u>Proof</u>. That $\{X_i\}$ exists and forms a process of the type asserted follows from the preceeding discussion and the fact that (6.2) implies (6.1). We proceed with the proof that Condition (i) of Theorem 3.1 holds, making use not of (6.2), but only of its consequence (6.1). We will use repeatedly the inequality (easily established by induction on n)

$$a_0^2 + (a_0 + a_1)^2 + \cdots + (a_0 + \cdots + a_{n-2})^2$$
$$+ \sum_{i=0}^{\infty} (a_i + a_{i+1} + \cdots + a_{i+n-1})^2 \leq n \left(\sum_{i=0}^{\infty} |a_i|\right)^2 .$$

Following Diananda, define

$$P_{i,k} = \sum_{t=0}^{k-1} A_t Y_{i-t} , \qquad Q_{i,k} = \sum_{t=k}^{\infty} A_t Y_{i-t} ,$$

$$U_{n,k,j} = n^{-1/2} \sum_{i=n_{j-1}+1}^{n_j} P_{i,k} ,$$

$$V_{n,k,j} = n^{-1/2} \sum_{i=n_{j-1}+1}^{n_j} Q_{i,k} .$$

Now for k fixed and i varying, $\{P_{i,k}\}$ is a stationary $(m+k-1)$-dependent process with zero means and finite variances. It was proved in the preceeding section that Condition (i) holds for such processes (Theorem 5.2). Therefore the distribution of

$$(6.4) \qquad (U_{n,k,1}, \ldots, U_{n,k,c})$$

approaches, as $n \to \infty$, the normal distribution having zero means and having as covariance matrix the identity multiplied by σ_k^2, where

$$(6.5) \qquad \sigma_k^2 = \lim_{n \to \infty} E\{U_{n,k,j}^2\} = c^{-1}(A_0 + \cdots + A_{k-1})^2 \lambda , \qquad j = 1, \ldots, c ,$$

where

$$(6.6) \qquad \lambda = E\{Y_0^2\} + 2 \sum_{v=1}^{m} E\{Y_0 Y_v\} .$$

To establish the second equality in (6.5) it is enough to show that

$$\lim_{n\to\infty} n^{-1} E\{(\sum_{i=1}^{n} P_{i,n})^2\} = (A_0 + \cdots + A_{k-1})^2 \lambda . \tag{6.7}$$

Let $\Sigma_{i=1}^{n} P_{i,k} = \xi + \eta$, where

$$\xi = (A_0 + \cdots + A_{k-1}) \sum_{v=1}^{n-k+1} Y_v ,$$

and

$$\eta = A_0 Y_n + (A_0 + A_1) Y_{n-1} + \cdots + (A_0 + \cdots + A_{k-2}) Y_{n-k+2}$$

$$+ (A_1 + \cdots + A_{k-1}) Y_0 + (A_2 + \cdots + A_{k-1}) Y_{-1} + \cdots + A_{k-1} Y_{2-k} .$$

Then

$$E\{\xi^2\} = (A_0 + \cdots + A_{k-1})^2 (\lambda (n-k+1) + \mu) ,$$

with λ defined by (6.6) and μ by

$$\mu = -2 \sum_{v=1}^{m} v E\{Y_0 Y_v\} . \tag{6.8}$$

If

$$A = \sup_{t \geq 0} |A_t| , \tag{6.9}$$

then

$$(6.10) \qquad E\{\eta^2\} \leq 2k^2 A^2 E\{Y_0^2\} .$$

By Schwarz' inequality,

$$(6.11) \qquad (E^{1/2}\{\xi^2\} - E^{1/2}\{\eta^2\})^2 \leq E\{(\xi+\eta)^2\}$$

$$\leq (E^{1/2}\{\xi^2\} + E^{1/2}\{\eta^2\})^2 .$$

By (6.10) and (6.11),

$$\left| [n^{-1} E\{(\sum_{i=1}^{n} P_{i,k})^2\}]^{1/2} - [n^{-1}(\lambda(n-k+1)+\mu)]^{1/2} \right|$$

$$\leq 2k^2 A^2 E\{Y_0^2\} n^{-1} ,$$

which yields (6.7), and hence (6.5). From the convergence of the distribution of (6.4), and the limit

$$\lim_{k \to \infty} \sigma_k^2 = c^{-1} \sigma^2 ,$$

where σ^2 is defined by (6.3), it follows that the distribution of

$$\sigma^{-1}(U_{n,k,1}, \ldots, U_{n,k,c})$$

approaches, if $n \to \infty$ and then $k \to \infty$, the normal distribution with zero means having as covariance matrix c^{-1} times the identity. In order to show that the distribution of

$$\sigma^{-1} n^{-1/2} (S_{n_1}, S_{n_2} - S_{n_1}, \ldots, S_{n_c} - S_{n_{c-1}})$$

converges, as $n \to \infty$, to this normal distribution, it suffices by Theorem A. 4 to prove that

$$\underset{k\to\infty}{\text{p lim}} \quad \underset{n\to\infty}{\text{p lim}} \quad (V_{n,k,1}, \ldots, V_{n,k,c}) = 0 \quad ,$$

or, by Chebyshev's inequality, that

$$\text{(6.12)} \qquad \lim_{k\to\infty} \limsup_{n\to\infty} E\{V^2_{n,k,j}\} = 0 \quad , \qquad j = 1, \ldots, c \quad .$$

Now, letting $w = n_j - n_{j-1}$, one computes

$$\text{(6.13)} \qquad \sum_{i=n_{j-1}+1}^{n_j} \sum_{t=k}^{\infty} A_t Y_{i-t} = A_k Y_{n_j} + (A_k + A_{k+1}) Y_{n_j - 1}$$

$$+ \cdots + (A_k + \cdots + A_{k+w-2}) Y_{n_{j-1}+2}$$

$$+ \sum_{v=0}^{\infty} (A_{k+v} + \cdots + A_{k+v+w-1}) Y_{n_{j-1}+1-v} \quad .$$

Letting

$$\text{(6.14)} \qquad B = \max_{0 \le v \le m} | E\{Y_0 Y_v\}| \quad ,$$

and using the well-known fact that $E\{Z^2\} \le \limsup_{n\to\infty} E\{Z_n^2\}$ if

$Z = p \lim_{n \to \infty} Z_n$, one deduces from (6.13) that

(6.15) $$E\{V_{n,k,j}^2\} \leq (n_j - n_{j-1})\, n^{-1} B(|A_k| + |A_{k+1}| + \cdots)^2,$$

which yields (6.12). This completes the proof that Condition (i) holds.

In order to prove that (3.3) holds in the present circumstance, we first show that there exists a sequence $\{\beta_n\}$ of integers tending to infinity in such a way that

(6.16) $$\lim_{n \to \infty} n\beta P\{|X_1| \geq \varepsilon \beta^{-1} n^{1/2}/2\} = 0$$

and

(6.17) $$\lim_{n \to \infty} \sum_{r=1}^{n} P\{|\sum_{v=0}^{\infty} Y_{-v}(A_{\beta+v} + \cdots + A_{n_{j,u+1}-r+v})| \geq \varepsilon n^{1/2}/4\} = 0.$$

It is easy to show that X_1 has a finite second moment. Hence there exists an increasing sequence $\{m_k\}$ such that

$$k\, n\, P\{|X_1| \geq \varepsilon k^{-1} n^{1/2}\} \leq k^{-1}$$

if $n \geq m_k$. If one puts $\beta_n = k$ for $m_k < n \leq m_{k+1}$, than β_n goes to infinity and (6.16) holds. It follows from (6.2) that

$$R_n = 0(n^{-2}) , \tag{6.18}$$

where $R_n = |A_n| + |A_{n+1}| + \cdots$. Now

$$P\{|\sum_{v=o}^{\infty} Y_{-v}(A_{\beta+v}+\cdots+A_{n_{j,u+1}-r+v})| \geq \varepsilon n^{1/2}/4\} \tag{6.19}$$

$$\leq P\{\sum_{v=o}^{\infty} |Y_{-v}| R_{\beta+v} \geq \varepsilon n^{1/2}/4\}$$

$$\leq \sum_{v=o}^{\infty} P\{|Y_{-v}| R_{\beta+v} \geq n^{1/2} (\beta+v)^{-5/4}\} ,$$

provided n is large enough that

$$\sum_{v=0}^{\infty} (\beta+v)^{-5/4} < \varepsilon/4 .$$

But

$$P\{|Y_{-v}| R_{\beta+v} \geq n^{1/2} (\beta+v)^{-5/4}\} \leq n^{-1} (\beta+v)^{5/2} R^2_{\beta+v} ,$$

so that by (6.18) the sum in (6.17) is dominated by $\Sigma_{v=0}^{\infty} (\beta+v)^{-3/2}$ which goes to zero since β goes to infinity. Hence (6.17).

We now decompose the summand in (3.3) into

$$P(E_{n,r} \cap \{|S_{n_{j,u+1}} - S_r| \geq \varepsilon n^{1/2}\}) \leq P\{|S_{r+\beta} - S_r| \geq \varepsilon n^{1/2}/2\} \tag{6.20}$$

$$+ P(E_{n,r} \cap \{|S_{n_{j,u+1}} - S_{r+\beta}| \geq \varepsilon n^{1/2}/2\}) ,$$

where $\{\beta_n\}$ satisfies (6.16) and (6.17). It follows immediately from stationarity and (6.16) that

$$(6.21) \qquad \lim_{n\to\infty} \sum_{r=1}^{n} P\{|S_{r+\beta} - S_r| \geq \varepsilon\, n^{1/2}/2\} = 0 .$$

Let $n_{j,u+1} - r = w$. Then $S_{n_{j,u+1}} - S_r = \xi + \eta$, where

$$\xi = \sum_{v=0}^{\infty} (A_{\beta+v} + \cdots + A_{w+v})\, Y_{r-v}$$

and

$$\eta = A_0 Y_{r+w} + (A_0 + A_1) Y_{r+w-1} + \cdots + (A_0 + \cdots + A_{w-\beta-1}) Y_{r+\beta+1}$$

$$+ \sum_{v=0}^{\beta-1} (A_v + \cdots + A_{v+w-\beta}) Y_{r+\beta-v} .$$

Hence

$$(6.22) \qquad P(E_{n,r} \cap \{|S_{n_{j,u+1}} - S_{r+\beta}| \geq \varepsilon\, n^{1/2}/2\})$$

$$\leq P(E_{n,r})\, P\{|\eta| \geq \varepsilon\, n^{1/2}/4\} + P\{|\xi| \geq \varepsilon\, n^{1/2}/4\} ,$$

where the factorization of the first term on the right is valid if n is large enough that $\beta_n > m$. Now by stationarity and (6.17) we have,

$$(6.23) \qquad \lim_{n\to\infty} \sum_{r=1}^{n} P\{|\xi| \geq \varepsilon\, n^{1/2}/4\} = 0 .$$

Using the m-dependence property, one computes

$$E\{\eta^2\} \le 2A^2 B m w + A_0^2 + (A_0 + A_1)^2 + \cdots + (A_0 + \cdots + A_{w-\beta-1})^2$$

$$+ \sum_{v=0}^{\beta-1} (A_v + \cdots + A_{v+w-\beta})^2$$

$$\le w\left(2A^2 B m + \left(\sum_{v=0}^{\infty} |A_v|\right)^2\right),$$

where A and B are defined by (6.9) and (6.14). Since, in the notation of Theorem 3.1, $w \le n/c\nu$,

$$\sum_{r=1}^{n} P(E_{n,r}) P\{|\eta| \ge \varepsilon n^{1/2}/4\} \le (\text{const.})\, 16 / \varepsilon^2 c\nu . \tag{6.24}$$

Finally, (3.3) follows from (6.20), (6.21), (6.22), (6.23) and (6.24), completing the proof of the theorem.

It is clear that (6.2) can be replaced by the weaker condition (6.18). In fact

$$R_n = 0(n^{-1-\delta}) \qquad , \qquad \delta > 0,$$

suffices. An examination of the proof shows that if there exists a sequence $\{\beta_n\}$ going to infinity in such a way that (6.16) and (6.17) hold, then the result follows. This fact can be used to weaken (6.2) under the assumption that Y_j possesses some moment of order higher than two.

§7. The invariance principle for recurrent events.

Let $\mathcal{E}$ be a recurrent event in the sense of Feller (cf. [11] or [12]). Suppose that $\mathcal{E}$ is certain, let

$$X_1, X_2, \ldots$$

be the successive recurrence times of $\mathcal{E}$ and let

$$S_k = X_1 + \cdots + X_k .$$

Let Z_n be 1 or 0 according as $\mathcal{E}$ occurs or not at the nth trial and let

$$N_n = Z_1 + \cdots + Z_n$$

be the number of occurrences of $\mathcal{E}$ during the first n trials. Assume that the recurrence times have finite mean μ and variance σ^2. In this section we prove the invariance principle for the sequence $\{Z_n - \mu^{-1}\}$.

Theorem 7.1. If the recurrence times have finite mean μ and variance σ^2, then the invariance principle holds for the sequence $\{Z_n - \mu^{-1}\}$ with norming factors $\sigma \mu^{-3/2} n^{1/2}$.

Proof: Feller has proved the central limit theorem for N_n by reducing it to the central limit theorem for S_k via the identity

$$(7.1) \qquad \{N_n \geq k\} = \{S_k \leq n\}.$$

Our proof that Condition (i) of Theorem 3.1 holds proceeds in the same way. Let $\Phi(a_1, \ldots, a_c)$ be the normal distribution with zero means and covariance matrix (λ_{ij}), where $\lambda_{ij} = c^{-1} \min(i, j)$. We must show that

$$(7.2) \qquad \lim_{n\to\infty} P\{N_{n_j} - n_j \mu^{-1} \leq a_j \sigma \mu^{-3/2} n^{1/2},\ j = 1, \ldots, c\}$$

$$= \Phi(a_1, \ldots, a_c).$$

For $j = 1, \ldots, c$ let $k_j = k_j(n)$ be one greater than the integral part of $n_j \mu^{-1} + a_j \sigma \mu^{-3/2} n^{1/2}$. Then (7.2) reduces to

$$(7.3) \qquad \lim_{n\to\infty} P\{N_{n_j} < k_j,\ j = 1, \ldots, c\} = \Phi(a_1, \ldots, a_c).$$

By (7.1) we see that (7.3) will follow if we can prove

$$\lim_{n\to\infty} P\left\{\frac{S_{k_j} - k_j \mu}{\sigma \mu^{-1/2} n^{1/2}} > \frac{n_j - k_j \mu}{\sigma \mu^{-1/2} n^{1/2}},\ j = 1, \ldots, c\right\} = \Phi(a_1, \ldots, a_c).$$

Since

$$n_j - k_j \sim -a_j \sigma \mu^{-1/2} n^{1/2}$$

and $\Phi(-a_1, \ldots, -a_c) = 1 - \Phi(a_1, \ldots, a_c)$, it suffices to show that the distribution of

$$\frac{1}{\sigma \mu^{-1/2} n^{1/2}} (S_{k_1} - k_1 \quad , \ldots, S_{k_c} - k_c \quad)$$

approaches Φ. But this follows easily from Theorems A.3 and A.4 and the fact that $k_j \sim n_j \mu^{-1}$.

We need a subsidiary result for the second part of the proof. For each integer r let U_r be the last trial before the $(r+1)$st at which $\mathcal{E}$ occurs, letting $U_r = 1$ if there is no occurrence of $\mathcal{E}$ in the first r trials. And let V_r be the first trial after the rth at which $\mathcal{E}$ occurs. Then

$$(7.4) \qquad P\{V_r - U_r = k\} = P\{X_1 = k\} .$$

For when $\mathcal{E}$ occurs the process begins anew (cf. [11]) so that

$$P\{V_r - U_r = k\} = \Sigma_{v=r-k+1}^{r} P\{U_r = v\} P\{X_1 = k\} = P\{X_1 = k\} .$$

To prove that Condition (ii) of Theorem 3.1 holds, define random variables $\beta_{n,r}$ by

$$\beta_{n,r} = \begin{cases} V_r & \text{if} \quad V_r \le n_{j,u+1} \\ n_{j,u+1} & \text{if} \quad V_r > n_{j,u+1} . \end{cases}$$

Then

$$(7.5) \qquad P(E_{n,r} \cap \{|(N_{n_{j,u+1}} - n_{j,u+1}\mu^{-1}) - (N_r - r\mu^{-1})| \ge \varepsilon n^{1/2}\})$$

$$\le P\{|(N_{\beta_{n,r}} - \beta_{n,r}\mu^{-1}) - (N_r - r\mu^{-1})| \ge \frac{\varepsilon}{2} n^{1/2}\}$$

$$+ P(E_{nr} \cap \{|(N_{n_{j,u+1}} - n_{j,u+1}\mu^{-1})$$

$$- (N_{\beta_{n,r}} - \beta_{n,r}\mu^{-1})| \ge \frac{\varepsilon}{2} n^{1/2}\}) .$$

Now $N_{\beta_{n,r}} - N_r$ is 0 or 1 (according as $V_r > n_{j,u+1}$ or not), so that $|(N_{\beta_{n,r}} - \beta_{nr}\mu^{-1}) - (N_r - \mu^{-1})| \le 1 + \mu^{-1}(\beta_{n,r} - r)$.

And $\beta_{n,r} - r \le V_r - U_r$, so that

$$P\{\beta_{n,r} - r \ge \frac{\varepsilon}{4} n^{1/2}\} \le P\{X_1 \ge \frac{\varepsilon}{4} n^{1/2}\} .$$

Hence if n is large enough that $\frac{\varepsilon}{4} n^{1/2} \ge 1$,

$$(7.6) \qquad \sum_{r=1}^{n} P\left\{ |(N_{\beta_{n,r}} - \beta_{n,r}\mu^{-1}) - (N_r - r\mu^{-1})| \geq \frac{\varepsilon}{2} n^{1/2} \right\}$$

$$\leq n P\left\{ X_1 \geq \frac{\varepsilon}{4} n^{1/2} \right\} \longrightarrow 0 \ ,$$

where the limit holds because X_1 has a finite second moment. On the other hand, by the defining properties of recurrent events, the second term of the second member of (7.5) is equal to

$$P(E_{nr})\, P\left\{ |N_{n_{j,u+1}-\beta} - (n_{j,u+1} - \beta)\mu^{-1}| \geq \frac{\varepsilon}{2} n^{1/2} \right\} \ .$$

It is shown in [11, p.111] that

$$E(N_k) = k\mu^{-1} + \frac{1}{2}(\sigma^2 + \mu + \mu^2)\mu^{-2} - 1 + o(1)$$

and

$$\mathrm{Var}\,\{N_k\} \sim k\sigma^2\mu^{-3} \ .$$

Hence there exists a constant A such that

$$E\{(N_k - k\mu^{-1})^2\} \leq Ak \ .$$

From this and Chebyshev's inequality it follows that

$$P\left\{ |N_{n_{j,u+1}-\beta} - (n_{j,u+1} - \beta)\mu^{-1}| \geq \frac{\varepsilon}{2} n^{1/2} \right\} \leq 4A/\varepsilon^2 c\nu.$$

Hence

$$(7.7) \qquad \lim_{\nu\to\infty} \limsup_{n\to\infty} \sum_{r=1}^{n} P(E_{nr} \cap \{|(N_{n_{j,u+1}} - n_{j,u+1}\mu^{-1}) - (N_\beta - \beta\mu^{-1})| \geq \frac{\varepsilon}{2} n^{1/2}\}) = 0 .$$

Finally, Condition (ii) follows from (7.5), (7.6) and (7.7), completing the proof of the theorem.

It is possible to prove this theorem by a direct extension of Feller's method, avoiding the use of Theorem 3.1. That is, it is possible to prove it using relation (7.1) and the invariance principle for $\{X_n - \mu\}$ (Theorem 5.3). While such a method is conceptually appealing, the details of the proof become very involved.

Appendix.

In this appendix we prove some multivariate central limit theorems and extend to several dimensions two theorems of Marsaglia [18]. The proofs are straightforward but seem not to be available in the literature in a generality sufficient for our purposes.

Let $\{V_{n,k},\ k=1, \ldots, \nu(n),\ n=1, 2, \ldots\}$ be an array of random vectors

$$V_{n,k} = (X_{n,k}^{(1)}, \ldots, X_{n,k}^{(c)}).$$

We assume that each $X_{n,k}^{(j)}$ has mean zero and a finite variance and that the vectors with a common first subscript are independent. Let $S_{n,j} = \Sigma_{k=1}^{\nu(n)} X_{n,k}^{(j)}$, $s_{n,j}^2 = E\{S_{n,j}^2\}$ and let A_n be the covariance matrix of the random vector

$$(s_{n,1}^{-1} S_{n,1}, \ldots, s_{n,c}^{-1} S_{n,c}). \tag{A.1}$$

In what follows we assume that

$$A = (a_{i,j}) = \lim_{n\to\infty} A_n \tag{A.2}$$

exists. The matrix A is of necessity positive semi-definite, and for covenience we assume it to be positive definite.

[illegible]

The first result says in effect that if (A.2) holds and the Lindeberg condition holds in each of the c components then the central limit theorem holds. The proof reduces the problem to the central limit theorem in one dimension, making use of a technique due to Cramér and Wold [4].

Theorem A.1. If (A.2) holds, where the limit matrix A is positive definite, and

$$\text{(A.3)} \qquad \lim_{n \to \infty} s_{n,j}^{-2} \sum_{k=1}^{\nu(n)} \int_{\{|X_{n,k}^{(j)}| \geq \varepsilon s_{n,j}\}} (X_{n,k}^{(j)})^2 \, dP = 0$$

for all $\varepsilon > 0$ and $j = 1, \ldots, c$, then the distribution of (A.1) converges to the normal distribution having A as its covariance matrix and zero means.

Proof. We must show that

$$\text{(A.4)} \qquad \lim_{n \to \infty} E\{\exp(i \sum_{j=1}^{c} t_j s_{n,j}^{-1} S_{n,j})\} = \exp(-1/2 \sum_{i,j=1}^{c} a_{i,j} t_i t_j)$$

for all $(t_1, \ldots, t_c)$. Let the t_j be arbitrary but fixed throughout the following discussion. For notational convenience we assume that no t_j is zero. Let

$$Y_{n,k} = \sum_{j=1}^{c} t_j s_{n,j}^{-1} X_{n,k}^{(j)} .$$

It is easily seen that the second moment of $\Sigma_{k=1}^{\nu(n)} Y_{n,k}$ is

$$s_n^2 = \sum_{i,j=1}^{c} a_{i,j}^{(n)} t_i t_j ,$$

where $a_{i,j}^{(n)}$ is the (i, j)th entry of A_n. Note that $s_n^2 > 0$ for sufficiently large n, since A is positive definite. If we can prove that the central limit theorem holds for the array $\{Y_{n,k}\}$ then

$$\text{(A.5)} \qquad \lim_{n\to\infty} E\,\{\exp(iu\, s_n^{-1} \sum_{k=1}^{\nu(n)} Y_{n,k})\} = e^{-u^2/2} .$$

Since

$$\text{(A.6)} \qquad \lim_{n\to\infty} s_n^2 = \sum_{i,j=1}^{c} a_{i,j}^{(n)} t_i t_j > 0$$

by (A.2), substituting $(\Sigma_{i,j=1}^{c} a_{i,j} t_i t_1)^{1/2}$ for u in (A.5) gives (A.4), in view of a well-known theorem (cf. Theorem A.5 below).

In order to show that the central limit theorem holds for the array $\{Y_{n,k}\}$ it is enough (cf., e.g., [13]) to show that

$$\text{(A.7)} \qquad \lim_{n\to\infty} s_n^{-2} \sum_{k=1}^{\nu(n)} \int_{\{|Y_{n,k}| \geq \varepsilon s_n\}} Y_{n,k}^2 \, dP = 0 .$$

for all $\varepsilon > 0$. Now choose ζ so that

$$0 < \zeta < c^{-1} \varepsilon \Big(\sum_{i,j=1}^{c} a_{i,j} t_i t_j \Big) \Big(\max_{1\leq j\leq c} |t_j| \Big)^{-1} ,$$

and let $\Lambda_{n,k}$ be the set where $|X_{n,k}^{(j)}| \geq \zeta s_{n,j}$ for some

$j = 1, \ldots, c$. By (A.6) and Hölder's inequality we have, for large n,

$$s_n^{-2} \sum_{k=1}^{\nu(n)} \int_{\{|Y_{n,k}| > \varepsilon s_n\}} Y_{n,k}^2 \, dP \leq c \, s_n^{-2} \sum_{k=1}^{\nu(n)} \sum_{j=1}^{c} \int_{\Lambda_{n,k}} s_{n,j}^{-2} (X_{n,k}^{(j)})^2 \, dP .$$

Using (A.6) again, we see that in order to prove (A. 7) it suffices to prove

$$\text{(A. 8)} \qquad \lim_{n \to \infty} s_{n,j}^{-2} \sum_{k=1}^{\nu(n)} \int_{\Lambda_{n,k}} (X_{n,k}^{(j)})^2 \, dP = 0$$

for $j = 1, \ldots, c$. But

$$s_{n,j}^{-2} \sum_{k=1}^{\nu(n)} \int_{\Lambda_{n,k}} (X_{n,k}^{(j)})^2 \, dP = s_{n,j}^{-2} \sum_{k=1}^{\nu(n)} \int_{\{|X_{n,k}^{(j)}| \geq \zeta s_{n,j}\}} (X_{n,k}^{(j)})^2 \, dP$$

$$+ s_{n,j}^{-2} \sum_{k=1}^{\nu(n)} \int_{\Lambda_{n,k} - \{|X_{n,k}^{(j)}| \geq \zeta s_{n,j}\}} (X_{n,k}^{(j)})^2 \, dP .$$

Now the first term of the right-hand member of the preceding inequality goes to zero by (A.3), while the second term is dominated by

$$\zeta^2 \sum_{k=1}^{\nu(n)} \int_{\Lambda_{n,k}} dP \;\leq\; \sum_{i=1}^{c} s_{n,i}^{-2} \sum_{k=1}^{\nu(n)} \int_{\{|X_{n,k}^{(i)}| \geq \zeta s_{n,i}\}} (X_{n,k}^{(i)})^2 \, dP \, ,$$

which goes to zero, again by (A.3). Thus we have proved (A.8) and hence the theorem.

The multivariate version of Lyapounov's theorem now follows immediately.

Theorem A.2. If (A.2) holds, where the limit matrix A is positive definite, if the $X_{n,k}^{(j)}$ all have finite moments of order $2+\delta$ for some $\delta > 0$, and if

$$\text{(A.9)} \qquad \lim_{n\to\infty} s_{n,j}^{-2-\delta} \sum_{k=1}^{\nu(n)} E\{|X_{n,k}^{(j)}|^{2+\delta}\} = 0$$

for $j = 1, \ldots, c$, then the distribution of (A.1) converges to the normal distribution having A as its covariance matrix and zero means.

Proof. Since

$$s_{n,j}^{-2} \int_{\{|X_{n,k}^{(j)}| \geq \varepsilon s_{n,j}\}} (X_{n,k}^{(j)})^2 \, dP \leq \varepsilon^{-\delta} s_{n,j}^{-2-\delta} E\{ |X_{n,k}^{(j)}|^{2+\delta} \} \, ,$$

(A.3) follows from (A.9).

A second consequence of Theorem A.1 is the multivariate central limit theorem for identically distributed random variables.

Theorem A.3. If the $V_{n,k}$ are identically distributed, each having the positive definite matrix A as its covariance matrix, and if $\lim_{n\to\infty} \nu(n) = \infty$, then the distribution of (A.1) converges to the normal distribution having A as its covariance matrix and zero means.

Proof. Let the distribution function of $X_{n,k}^{(j)}$ be F_j, where $\int_{-\infty}^{\infty} x^2\, dF_j(x) = \sigma_j^2$. Then

$$s_{n,j}^{-2} \sum_{k=1}^{\nu(n)} \int_{\{|X_{n,k}^{(j)}| \geq \varepsilon s_{n,j}\}} (X_{n,k}^{(j)})^2\, dP = \sigma_j^{-2} \int_{\{|x| \geq \varepsilon \sigma_j \nu^{1/2}(n)\}} x^2\, dF_j(x)$$

goes to zero, and the result follows.

The next two theorems are trivial extensions to c dimensions of results of Marsaglia [18] on iterated limits. Here an iterated limit is used in the strong sense, i.e., $\lim_{k\to\infty} \lim_{n\to\infty} a_{n,k} = a$ means $\lim_{k\to\infty} \limsup_{n\to\infty} |a - a_{n,k}| = 0$. If $V = (V^{(1)}, \ldots, V^{(c)})$ and $V_{n,k} = (V_{n,k}^{(1)}, \ldots, V_{n,k}^{(j)})$ are c-dimensional random vectors then by

$$\operatorname*{p\,lim}_{k\to\infty} \operatorname*{p\,lim}_{n\to\infty} V_{n,k} = V$$

we mean

$$\lim_{k\to\infty} \lim_{n\to\infty} P\{|V_{n,k}^{(j)} - V^{(j)}| < \varepsilon, j = 1, \ldots, c\} = 1$$

for all $\varepsilon > 0$.

<u>Theorem A.4.</u> Let $V_{n,k} = (V_{n,k}^{(1)}, \ldots, V_{n,k}^{(c)})$ and $U_{n,k} = (U_{n,k}^{(1)}, \ldots, U_{n,k}^{(c)})$ be random vectors such that

$$\operatorname*{p\,lim}_{k\to\infty} \operatorname*{p\,lim}_{n\to\infty} U_{n,k} = 0$$

and

$$\lim_{k\to\infty} \lim_{n\to\infty} P\{V_{n,k}^{(j)} \leq x_j,\ j = 1, \ldots, c\} = G(x_1, \ldots, x_c)$$

for all continuity points $(x_1, \ldots, x_c)$ of a distribution function G. Then

$$\lim_{k\to\infty} \lim_{n\to\infty} P\{V_{n,k}^{(j)} + U_{n,k}^{(j)} \leq x_n,\ j=1, \ldots, c\} = G(x_1, \ldots, x_c)$$

at continuity points of G.

<u>Proof.</u> Marsaglia's proof goes over almost word for word. Let $\varepsilon = 8\delta > 0$ and a continuity point $(x_1, \ldots, x_c)$ of G be given. We must find integers $K, N_1, N_2, \ldots$ such that

$$\text{(A.10)} \qquad |P\{V_{n,k}^{(j)} + U_{n,k}^{(j)} \leq x_j,\ j=1, \ldots, c\} - G(x_1, \ldots, x_c)| < \varepsilon$$

if $k > K$ and $n > N_k$. First choose β so that $(x_1 - \beta, \ldots, x_c - \beta)$ and $(x_1, +\beta, \ldots, x_c + \beta)$ are continuity points of G and so that

$$\text{(A.11)} \qquad 0 \leq G(x_1 + \beta, \ldots, x_c + \beta) - G(x_1 - \beta, \ldots, x_c - \beta) < \delta \quad .$$

Next choose $K, N_1, N_2, \ldots$ so that

(A.12) $$P\left(\bigcup_{j=1}^{c} \{ |U_{n,k}^{(j)}| > \beta \} \right) < \delta ,$$

(A.13) $$| P\{V_{n,k}^{(j)} \leq x_j, \ j=1, \ldots, c\} - G(x_1, \ldots, x_c) | < \delta ,$$

(A.14) $$| P\{V_{n,k}^{(j)} \leq x_j - \beta, \ j=1, \ldots, c\} - G(x_1 - \beta, \ldots, x_c - \beta) | < \delta ,$$

(A.15) $$| P\{V_{n,k}^{(j)} \leq x_j + \beta, \ i=1, \ldots, c\} - G(x_1 + \beta, \ldots, x_c + \beta) | < \delta ,$$

provided $k > K$ and $n > N_k$. For such a pair (n, k) set

$$F(\xi) = P\{V_{n,k}^{(j)} + U_{n,k}^{(j)} \leq x_j + \xi, \ j=1, \ldots, c\},$$

$$H(\xi) = P\{V_{n,k}^{(j)} + U_{n,k}^{(j)} \leq x_j + \xi, \ |U_{n,k}^{(j)}| \leq \beta, \ j=1, \ldots, c\},$$

$$L(\xi) = P\{V_{n,k}^{(j)} \leq x_j + \xi, \ |U_{n,k}^{(j)}| \leq \beta, \ j=1, \ldots, c\},$$

$$Q(\xi) = P\{V_{n,k}^{(j)} \leq x_j + \xi, \ j=1, \ldots, c\},$$

$$T(\xi) = G(x_1 + \xi, \ldots, x_c + \xi).$$

Then

$$|F(0) - T(0)| \leq |F(0) - H(0)| + |H(0) - L(0)| + |L(0) - Q(0)| + |Q(0) - T(0)| .$$

The first, third and fourth terms on the right are each less than δ

by (A.12), (A.12) and (A.13) respectively. Since

$$L(-\beta) \leq H(0) \leq L(\beta)$$

and

$$L(-\beta) \leq L(0) \leq L(\beta),$$

we have

$$|H(0) - L(0)| \leq |L(\beta) - L(-\beta)| \leq |L(\beta) - Q(\beta)|$$

$$+ |Q(\beta) - T(\beta)| + |T(\beta) - T(-\beta)| + |T(-\beta) - Q(-\beta)|$$

$$+ |Q(-\beta) - L(-\beta)| < 5\,\delta$$

by (A.12), (A.15), (A.11), (A.14) and (A.12). Hence $|F(0) - T(0)| < 8\,\delta$, which is (A.10).

Theorem A.5. Let $V_{n,k}$ be random vectors such that

$$\lim_{k\to\infty} \lim_{n\to\infty} P\{V_{n,k}^{(j)} \leq x_j,\ j=1, \ldots, c\} = G(x_1, \ldots, x_c)$$

at continuity points of G. Suppose that

$$\lim_{k\to\infty} \lim_{n\to\infty} \lambda_{n,k}^{(j)} = 1$$

for $j=1, \ldots, c$. Then

$$\lim_{k\to\infty} \lim_{n\to\infty} P\{\lambda_{n,k}^{(j)}\, V_{n,k}^{(j)} \leq x_j,\ j=1, \ldots, c\} = G(x_1, \ldots, x_c)$$

at continuity points of G.

<u>Proof.</u> By the preceding result it suffices to show that

$$\underset{k \to \infty}{\text{p lim}} \ \underset{n \to \infty}{\text{p lim}} \ ((\lambda_{n,k}^{(1)} - 1)V_{n,k}^{(1)}, \ldots, (\lambda_{n,k}^{(c)} - 1)V_{n,k}^{(c)}) = 0 ,$$

which is easy.

Bibliography.

[1] P. Alexandroff and H. Hopf, Topologie, Berlin, 1935.

[2] S. Bernstein, Sur l'extension du théorème limite du calcul des probabilités aux sommes de quantités dépendants, Math. Ann., vol. 97 (1927), pp. 1-59.

[3] H. Cramér, Mathematical methods of statistics, Princeton, 1946.

[4] H. Cramér and H. Wold, Some theorems on distribution functions, J. London Math. Soc., vol. 11 (1936), pp. 290-295.

[5] P. Diananda, Some probability limit theorems with statistical applications, Proc. Cambridge Philos. Soc., vol. 49 (1953), pp. 239-246.

[6] M. Donsker, An invariance principle for certain probability limit theorems, Mem. Am. Math. Soc., no. 6, (1951).

[7] J. L. Doob, Heuristic approach to the Kolmogorov-Smirnov theorems, Ann. Math. Stat., vol. 20 (1949), pp. 393-403.

[8] J. L. Doob, Stochastic processes, New York, 1953.

[9] P. Erdös and M. Kac, On certain limit theorems in the theory of probability, Bull. Am. Math. Soc., vol. 52 (1946), pp. 292-302.

[10] P. Erdös and M. Kac, On the number of positive sums of independent random variables, Bull. Am. Math. Soc., vol. 53 (1947), pp. 1011-1020.

[11] W. Feller, Fluctuation theory of recurrent events, Trans. Am. Math. Soc., vol. 67 (1949), pp. 98-119.

[12] W. Feller, An introduction to probability theory and its applications, New York, 1950.

[13] B. V. Gnedendo and A. N. Kolmogorov, Limit distributions for sums of independent random variables, Cambridge, 1954.

[14] W. Hoeffding and H. E. Robbins, The central limit theorem for dependent random variables, Duke Math. J., vol. 15 (1948), pp. 773-780.

[15] A. N. Kolmogorov, Grundbegriffe der Wahrscheinlichkeitsrechnung, Berlin, 1933.

[16] P. Lévy, Théorie de l'addition des variables aléatoires,

[17] A. M. Mark, Some probability limit theorems, Bull. Am. Math. Soc., vol. 55 (1949), pp. 885-900.

[18] G. Marsaglia, Iterated limits and the central limit theorem for dependent variables, Proc. Am. Math. Soc., vol. 6 (1954), pp. 987-991.

[19] Yu. V. Prohorov, Probability distributions in functional spaces, Uspehi Matem. Nauk (N.S.) 8, no. 3(55) (1953), pp. 165-167. (Russian.)

Thesis 28883

B542 Billingsley

The invariance principle for dependent random variables.